**Univerzita Karlova**

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Diplomová práce

Barbora Bulantová

# Syntactic Complexity in the Speech of Learners of English: Issues in Operationalization

Syntaktická komplexnost anglického jazyka nerodilých mluvčích a její operacionalizace

Praha 2020

PhDr. Tomáš Gráf, Ph.D.

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 26.05.2020 Barbora Bulantová

Tímto bych chtěla velmi poděkovat vedoucímu své práce, PhDr. Tomáši Gráfovi, PhD., za jeho nekonečnou podporu, ochotu a cenné rady.

# Abstrakt

Práce analyzuje produktivní a strukturní syntaktickou komplexnost monologů vyprodukovaných dvaceti nerodilými mluvčími angličtiny na úrovních B2 a C1. Data pocházejí z LINDSEI_CZ (Gráf 2017). Přepisy nahrávek byly rozděleny do AS-units a upraveny pro analýzu syntaktické komplexnosti podle instrukce Foster et al. (2000). Zhodnoceny jsou zároveň problematické případy v analýze mluvených dat, které Foster et al. (2000) neuvádějí. Produktivní syntaktická komplexnost byla změřena pomocí následujících metrik: průměrná délka AS-unit, průměrná délka klauze, počet klauzí / AS-unit. Metriky produktivní komplexnosti byly doplněny specifickými indikátory strukturní komplexnosti, kterými byly poměry výskytů několika druhů vedlejších vět a souřadných slovesných frází k celkovém počtu klauzí (Vercellotti & Packer 2016) a vážená škála strukturní komplexnosti navržená Vercellotti (2018).

Výsledky kvantitativní analýzy neprokázaly signifikantní vliv jazykové úrovně na syntaktickou komplexnost měřených monologů. Tato práce tedy doporučuje začlenění specifičtějších strukturních indikátorů do výzkumu mluvené komplexnosti. Variace ve skórech mezi řečníky, která byla v práci zjištěna, naznačuje, že důležitou roli ve výzkumu hraje idiosynkrasie. Výstup této diplomové práce je zároveň metodický, jelikož jsou identifikovány kritické oblasti, kterými by se měl výzkum komplexnosti cizího jazyka více zabývat.



**Klíčová slova**: CAF, syntaktická komplexnost, AS-unit, mluvený jazyk, žákovský jazyk



# Abstract

The thesis analyses syntactic complexity of monologic tasks of 10 B2 and 10 C1 speakers of English with Czech as their L1. The data derives from LINDSEI_CZ (Gráf 2017). The transcripts of the recordings were segmented into AS-units (Foster et al. 2000) and adapted for the purposes of the analysis. Syntactic complexity was calculated using following measures: mean length of AS-unit, mean length of clause, clauses / AS-unit. These were complemented by fine-grained indices of structural complexity, comprised of ratios of subordinate clause types and coordinate verb phrases per total number of clauses (Vercellotti & Packer 2016) and a weighted complexity scale designed by Vercellotti (2018).

The results of the quantitative analysis showed no significant effect of proficiency on syntactic complexity of the speakers. In fact, all speakers irrespective of language proficiency level produced very similar complex structures as opposed to lower-proficiency structures. This indicates that more fine-grained indices should be incorporated into spoken complexity research. At the same time, scores of each measure varied considerably within groups, pointing at the importance of inter-speaker variability in this research. The thesis thus produced outcomes that are largely methodological, in that it has identified critical areas which need to be scrutinized in L2 complexity research.



**Keywords**: CAF, syntactic complexity, AS-unit, spoken language, learner language

# Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| **CAF** | complexity, accuracy, fluency |
| **C/AS** | clauses / AS-units |
| **CEFR** | Common European Framework of Reference |
| **CHAT** | Codes for Human Analysis Transcript |
| **EFL** | English as a foreign language |
| **ICLE** | International Corpus of Learner English |
| **L1** | first language |
| **L2** | second language |
| **L2SCA** | L2 Syntactic Complexity Analyser |
| **LINDSEI** | Louvain International Database of Spoken English Interlanguage |
| **MLC** | mean length of clause |
| **MLU** | mean length of unit |
| **SALT** | Systematic Analysis of Language Transcripts |
| **SLA** | second language acquisition |
| **TAASC** | Tool for the Automatic Analysis of Syntactic Complexity |

# 0 Introduction

The study of L2 proficiency has been a major topic of interest for more than a century, and the endeavours to determine its constituents have spanned across various areas of research in applied linguistics. Until now, a number of dominant language proficiency models has arisen, among these e.g. established four-skills model and other influential cognitive and sociolinguistic models (Canale and Swain 1980, Kramsch 1986, Bachman 1990, Bialystok 1994 to name a few). Along these, the Complexity, Accuracy, and Fluency model (henceforth CAF) (see e.g. Housen et al. 2012) has proved to be their apt counterpart in measuring language performance.

Complexity, perhaps the most complex component of the triad, is typically operationalized with respect to several levels of language performance, ranging from syntactic and lexical to morphological and phonological. In terms of syntactic complexity, however, the research has largely focused on the written mode of production, while spoken language has been paid only limited attention. There exist several highly accurate automated tools for analysis of written syntactic complexity (e.g. McNamara et al. 2010, Lu 2010, Kyle 2016), whereas automation of spoken-language-complexity analysis with as accurate results as those of above-mentioned tools is debatable, to say the least. This is partly due to the lack of unified methodology of text segmentation and data exclusion, which should reflect the specifics of spoken language and the component of the CAF model analysed. When segmenting transcribed data, researchers often avail themselves of mere definitions of syntactic units and choose to omit their decisions in disputable cases. The tangled reality of spoken language requires a multi-faceted approach to data segmentation and exclusion, the lack of which entails different findings and hinders emerging of automated tools designed solely for spoken texts.

The aim of the thesis is to determine whether there is a significant difference in syntactic complexity of informal spoken production in s total of twenty B2 and C1 EFL learners with Czech

as their L1. The dataset derives from the Czech counterpart of the LINDSEI corpus (Gráf 2017). To ensure a comprehensive approach to measuring syntactic complexity, the study opts for measures of overall syntactic, sentential, and supra-sentential complexity as listed in Norris and Ortega (2009) and Housen (2012) and fine-grained structural indices designed by Vercellotti & Packer (2016) and Vercellotti (2018). As there are several automated tools available to compute these measures in written texts, a subsidiary aim of the study is to discover whether computation of syntactic complexity of spoken texts can be similarly automated using these tools.

There should be two intended outcomes of the thesis. It should firstly contribute to spoken-language-complexity research by analysing oral tasks of advanced students of English with Czech as their first language. The other outcome is purely methodological, in that the thesis should identify areas in the research of spoken syntactic complexity that decrease validity and replicability of the studies and suggest relevant solutions that might resolve such issues.

The thesis consists of 6 chapters. Chapter 1 introduces the concepts of complexity, accuracy, and fluency and describes their development into an established L2 proficiency model. Chapter 2 focuses solely on complexity and its operationalization. It also offers an overview of methodologies of text segmentation and data exclusion and includes a section pertaining to automated tools for analysing syntactic complexity of written texts. Chapter 3 presents the research questions. The final three chapters constitute the empirical part of the thesis. Chapter 4 describes the data set and LINDSEI_CZ. Chapter 5 describes the process of data segmentation, data exclusion, and the measures used to obtain the results. Finally, Chapter 6 presents the results and the discussion.

# 1 Language Proficiency and The CAF Model

The question of what it means to be a proficient language user has been at the heart of any research field that is in any way associated to language learning and language acquisition. The very definition of language proficiency and its categorization is something that helps this

vast, vague enterprise materialize, which in practical terms entails identifying valid research variables and institutionalization of language skills. On the level of European language education, institutions ranging from public schools to language agencies to specialized language centres have mostly relied on the Common European Framework Reference (CEFR), which defines each level of proficiency in terms of so-called *can-do statements* (Council of Europe 2001). In the context of language classroom, the practice is typically formed on the grounds of the four-skills model (the skills being reading, writing, listening, speaking). Many researchers have proposed, however, that in order to determine language proficiency, it is vital to measure its direct product – written and spoken production (Michel 2017). In practice, language skills are most frequently (and, perhaps, most importantly) demonstrated in interaction with other speakers.

After the decades of language learning and acquisition research, there is now a general consensus that language performance is, just as proficiency, inherently multicomponential (Housen et al. 2012). The CAF triad emerged as a new research opportunity to gauge language performance in its entirety and encompass all its dimensions. The interest in the three components can be traced back into the 1970s (Housen et al. 2012, Wolfe-Quintero et al. 1998), but it was in the context of communicative classroom research that they came to be studied together. Brumfit (1984) based his teaching model on the distinction between fluency-focused and accuracy-focused teaching strategies and activities with the aim to provide learners with an effective preparation for communicative contexts outside the classroom. He proposes their separate practice on the grounds that the focus on one of them might inhibit the production in terms of the other. In the 1990s Skehan (1998) added complexity and thus gave rise to a new proficiency model, which successfully complemented the body of already established models. Larsen-Freeman (1977, 1978), in contrast, approached language proficiency from a different perspective, drawing from L1 research in an attempt to design an index of development

applicable to L2 context, which would eventually lead to the identification of individual developmental stages of second language acquisition (Hakuta 1975). The index, although controversial, operated on the basis of a measure similar to Mean Length of Unit (MLU) (Michel 2017) (for more on MLU see section 2.2.1.3).

The CAF model has gained significant influence in L2 research and has appeared in countless studies, with the three components most commonly treated as dependent variables. This means that changes in complexity, accuracy, and fluency were studied with the aim to observe other variables in the research (among others task design, planning, individual differences, and differences in instruction) (Skehan & Foster 1997, Yuan & Ellis 2003, Housen et al. 2012, Ellis & Barkhuizen 2005). The findings pointed at various patterns of interaction between individual components of the CAF model. Skehan (1998) and Robinson (2001) were the strongest voices to purport that parameters of language performance may vary depending on the goal that the learner wants to achieve. Each proposed, however, a different theory as to the source of their interaction. Skehan (1998) in his Limited Attentional Capacity Model claims that a trade-off effect, i.e. an increase in one area of performance resulting in a decrease in another, appears due to limitations in attentional resources and working memory. If the learner chooses to focus on error-free production, the competition for attentional resources will likely entail a decrease in complexity. An effect similarly notable to the interaction of accuracy and complexity was observed also with accuracy and fluency. In a theory known as Cognition Hypothesis Robinson, Robinson (2001) asserts that it is cognitive and functional demands of the task that encourage the speaker to produce more accurate and complex language. Neither of the hypotheses, however, produced robust empirical evidence (Robinson 2011, Housen et al. 2012, Michel 2017).

Some more recent studies have stressed the importance the of interaction of complexity, accuracy, and fluency in longitudinal research. Larsen-Freeman (2009) pointed out that the

language development in terms of the three components is not collinear, and that the changes in their interaction in time should also be investigated. That it is important not to examine them one by one but treat them as a functioning whole has also been asserted in connection with their validity as markers of proficiency, as factor analyses have shown that complexity, accuracy, and fluency are "distinct and competing areas of L2 performance" (Housen et al. 2012, p. 3), which entails that only together can they serve as a basis for general conclusions about language performance and proficiency (ibid).

## 1.1 CAF: Challenges & Constraints

After years of complexity, accuracy, and fluency research, we are now faced with several deficiencies of the model with regard to definitions of the constructs, operationalization of CAF, and addressing other factors of language performance, such as communicative aims.

### 1.1.1 Defining CAF

Despite the high quantity of studies that has investigated various effects on complexity, accuracy, and fluency, the working definitions of the concepts have neither changed, nor been questioned. Although CAF did come under scrutiny in various volumes (e.g. Ellis & Barkhuizen 2005, Pallotti 2009, Housen et al. 2012, Michel 2017), many researchers have failed to provide a clear and sufficient definition of what is actually measured, or have only worked with vague descriptions of the concepts (Housen et al. 2012). A reason for this persisting controversy might be the lack of theory that the concepts would constitute (Pallotti 2009). A perhaps more tangible reason is the fact that the concepts themselves are as multidimensional as language proficiency and performance. Fluency is multicomponential in nature (Tavakoli & Skehan 2005), and all of its basic definitions presuppose a norm of a native speaker (Pallotti 2009). Accuracy, has been questioned as a valid indicator of language development (see e.g. Pallotti 2009), while

complexity is a conundrum of overlapping definitions and categories (for more on complexity, see section 2.1).

### 1.1.2 Measure Validity

A frequent phenomenon known to inhibit replicability of many studies and validity of their research is the quantity and the dubious quality of the measures that they employ. The first problem, i.e. the enormous range of measures available for each of the triad (for lists of measures see e.g. Wolfe-Quintero et al. 1998, Ellis & Barkhuizen 2005, or Housen et al. 2012), is an aggravating factor for the replicability of studies and comparison of findings. A greater concern has been voiced towards the quality of measures, i.e. whether they measure what they ought to. Many studies have employed a wide scale of metrics, several of which pertained to very similar phenomena (Norris & Ortega 2009, Housen et al. 2012). Pallotti (2009) further criticizes researchers who tend to conclude that the best measures are those that would show variation among the subjects, while ignoring the fact that similarity might also point to valid results.

### 1.1.3 Communicative Adequacy

Finally, a serious limitation of the CAF model is its focus on purely linguistic outcomes, while it has completely disregarded communicative adequacy of language performance. Most speakers learn a language with the aim to communicate successfully – this is a fact that has been at heart of modern language classroom practice and communicative language teaching. Pallotti (2009) aptly remarks that a speaker may produce a text which scores high in complexity, accuracy, and fluency without "getting the message across" and vice versa. Although the communicative function has not yet been made a priority in the CAF research, several studies have emerged which examined complexity, accuracy, and fluency alongside communicative adequacy (e.g. Kuiken et al. 2010, Révész et al. 2014).

# 2 Syntactic Complexity and Its Operationalization

As the focal point of the thesis is syntactic complexity of spoken learner language, this section does not provide any further information on accuracy and fluency. This chapter introduces the construct of complexity and all its dimensions. It then focuses on syntactic complexity and the constrains in its operationalization. Another concern of this chapter is the analysis of spoken language, particularly the lack of method on segmentation of transcribed data into units, and on exclusion of words that could skew the results.

## 2.1 L2 Complexity

In almost every account written on CAF since the 1990s, complexity has been introduced as the notoriously complex and controversial constituent of the triad, which is in addition difficult to operationalize (e.g. Housen & Kuiken 2009, Norris & Ortega 2009, Pallotti 2009, Housen et al. 2012, Pallotti 2015, Michel 2017). The biggest issue in L2 research so far has been the polysemous nature of the theoretical construct, which many studies failed to sufficiently answer and clarify in their definitions (Pallotti 2015). Housen et al. (2012) see either too vague or entirely missing definitions as causation of contradictory results of such studies, which they exemplify with following definitions:

- "[complexity is the] use of more challenging and difficult language … Complexity is the extent to which learners produce elaborated language" (R. Ellis & Barkhuizen 2005: 139)
- "Grammatical and lexical complexity mean that a wide variety of both basic and sophisticated structures and words are available to the learner" (Wolfe-Quintero, Inagaki, & Kim 1998: 69, 101)
- "Complexity refers to … the complexity of the underlying interlanguage system developed" (Skehan 2003: 8)

But the problem of polysemy is rooted at a much more basic level of description. Cambridge Dictionary defines complexity as "the state of having many parts and being difficult to understand or find an answer to" (Cambridge University Press 2020). There are two distinct senses that the definition distinguishes, one objective, generic quality of systems and elements in them, and the other, perceived sense, pertaining to human cognition. Theoretical perspectives on complexity operate on similar grounds. Rescher (1998) sees complexity as "a matter of the quantity and variety of its constituting elements and of interrelational elaborateness of their organizational and operational make-up" (p. 1). It is an objective, independent property of systems in the real world, but it is the limited human cognition and the process of understanding it that is dependent on the complexity models and hierarchies (ibid.). Miestramo (2008), describing different languages from the typological point of view, established two approaches to complexity based on this contrast, where the *relative* approach is entirely subjective and relates to the perceived difficulty to understand a language system or a language item. The *absolute* approach, on the contrary, has to do with properties of language systems.[1]

Boulté & Housen (2012) take up the *relative*/*absolute* approach and elaborate on it in their "descriptive analytic framework" (p. 26), a research-based taxonomy of complexity constructs, which should serve as a basis for future L2 complexity studies. They define the basic-level complexity from the *absolute* point of view as the "the number and the nature of the discrete components that the entity consists of, and […] the number and the nature of the relationships between the constituent components" (ibid., p. 22). Boulté & Housen (2012) further argue that there are at least three subdimensions within the L2 complexity framework: *propositional* complexity, *discourse-interactional* complexity and *linguistic* complexity. In L2

[1] The basic idea behind the approach is that the more elements there are in a language system, the more complex it becomes. (Miestramo 2008)

literature, the first two have received only marginal attention compared with the last dimension. Propositional complexity refers to a number of idea units encoded in one written or spoken message (the higher number of idea units, the more propositionally complex the message). Propositions to analyse discourse-interactional complexity have appeared mainly in studies on dialogic discourse, where it pertains to how many turns the speaker took and what kinds of turns they were.

It is only the linguistic (or *system, global*) complexity that is dealt with in more detail in the volume, where it is referred to as the "size, elaborateness, richness, and diversity of the learner's linguistic L2 system" (Housen & Kuiken 2009). It is the extent to which a learner masters complex language structures and is able to work with advanced and varied vocabulary in both written or spoken production. Linguistic complexity is manifested on various levels of language performance (phonology, lexis, morphology, syntax). When analysing learner language, we can either examine the global dimension of complexity (discussed above), or follow structural dimension of linguistic complexity, which relates to depth and breadth of individual linguistic structures in the learner's language repertoire (Boulté & Housen 2012). Each respective level then includes further specific subdimensions (clause, phrase, inflection, derivation…).

Even though this taxonomy might shed light on multidimensionality of complexity constructs in theory, in practice the dimensions are interrelated and usually appear together in research. Boulté & Housen (ibid.) warn that their taxonomy should not be considered as a complexity theory, as it was constructed to reflect the research published in recent years. They call for a theory that would unify now contradictory approaches to individual language features in terms of their complexity. Another concern of theirs is that in many studies, the underlying complexity construct only becomes apparent through specific quantitative measures that are employed (i.e. on the level of operationalization), while the theoretical dimension is somewhat

neglected. Three years after Housen et al. (2012), Pallotti (2015) revisited the issue and came to similarly inauspicious conclusions despite the advances made by his predecessors. Consequently, his article proposes a simple, purely descriptive structural view of linguistic complexity, where *system* complexity represents the *langue*, i.e. the complexity of the whole language system, and *text* complexity refers to "complexity of a given piece of discourse", i.e. *parole* (p. 120) (more on their operationalization in section 2.2.1).

### 2.1.1 Limitations of the Construct

Conceptually, linguistic complexity is not without its problems. Several factors should be addressed as to validity of complexity as an indicator of language proficiency. In this respect, Michel (2017) draws our attention to a general tendency to assume that a highly linguistically complex text signifies a higher level of L2 use. This is a fallacy for several reasons:

(1) It is a deliberate decision that stands behind (syntactically or lexically) complex language. Skehan (1998) stresses that a speaker is motivated by the circumstances (e.g. the task, communicative intention) to opt for more complex language items, the negative side effects of which are the trade-off effects between complexity and accuracy. It is therefore disputable that complexity could indicate a level of proficiency in spontaneous informal speech.

(2) Pallotti (2015) distinguishes between grammatical and so called "stylistic" syntactic complexity, which is a degree of stylistic variation within syntactic pattern caused by "culture-specific rhetorical patterns" (p. 120) and may vary based on individual stylistic choices and preferences.

(3) Stylistic variation brings us to the notion of genre. Michel (2017) gives an apt example of various writers, whose works all have an enormous literary value, but cannot be judged in terms of their syntactic complexity, which is, to a large extent, again a stylistic

choice. In some genres, lexically and syntactically complex language is one of its key norms (e.g. academic prose). In others, using overtly complex language could appear inappropriate, such as in informal conversations, which are in many cases pre-defined by social conventions. We should thus consider differences between written and spoken language, which are inherently very different and marked by distinct purposes. Spoken language, and informal spontaneous speech in particular, is by nature less complex. As Miller & Weinert (2009) prove, syntactic structure of phrases and clauses in spoken language differs from written language in its (lower) complexity.

(4) Following the previous argument, in spontaneous spoken language, speakers are only able to operate with smaller chunks of information at a time due to their limited processing capacity (Chafe 1987, Miller & Weinert 2009). This entails that even though a speaker focuses on syntactic complexity in spontaneous spoken language, the grammatical complexity is still affected by limitations in short-term memory, which bars more than one or two information units to be processed at a single time (Miller & Weinert 2009). In other words, spoken language is inherently extremely cognitively demanding and it is thus a misconception to consider spoken complexity as a marker of overall language proficiency.

## 2.2 Syntactic Complexity

Syntactic complexity is arguably the most frequently studied branch in linguistic complexity analysis (De Clercq & Housen 2017). Following the terminology and conceptual framework of Housen et al. (2012) and Ellis & Barkhuizen (2005), syntactic complexity is defined in this thesis as elaborateness and sophistication of language production in terms of its syntax, i.e. the number and the nature of discrete syntactic structures that a speaker produces, and the number and the nature of relationships between them and their constituting elements.

Many researchers have associated syntactic complexity with the length of various structures, which is why length-based measures regularly appear as the only indicator of syntactic complexity (De Clercq & Housen 2017) (more on syntactic complexity measures in section 2.2.1.2).

Studies observing development in syntactic complexity of learner language have largely been informed by the research in L1 language development, and it is a frequent assumption that acquisition of syntactic structures in L2 develops along the same lines as acquisition of syntactic structures in the mother tongue and thus progresses from simple to more elaborate and complex structures (see e.g. Norris & Ortega 2009, de Clercq & Housen 2010, for L1 development see Halliday and Mattheisen 1999). In earlier stages of development, learners start juxtaposing ideas via coordination, and identify basic relationships between them. Subsequently, they are able to assign logical relations between the ideas through hypotaxis. Finally, in advanced stages, they recognize grammatical metaphor and make use of nominalizations, which results in reduced production of subordinate clauses and complexification of phrasal structures (Norris & Ortega 2009, Halliday & Mattheisen 1999). Based on this line of development, coordination should be the most predictive in lower proficiency levels, the degree of subordination as opposed to coordination gains its importance in analysing intermediate level language production, while phrasal complexity proves as a highly useful indicator of syntactic complexity in upper-intermediate to proficient learners (Norris & Ortega 2009). This assumption of linear development has been challenged by Housen et al. (2012), who advocate the approach to syntactic complexity as a multidimensional notion.

### 2.2.1 Operationalizing Syntactic Complexity

Syntactic complexity should be operationalized on many more levels than those of coordination, subordination, and phrasal complexity, and using only length-based measures

appears to be rather reductionist. Depending on the research questions, the choice of appropriate measures enables us to discover general trends in longitudinal language development or detect finer inter-learner variation across all proficiency levels. The form of individual measures is then largely dependent on the chosen unit of analysis. Units can range from phrases and clauses to supra-clausal units. A problem arises in relation to task modality, i.e. in the distinction between written and spoken language. Modern schools of linguistics believed in primacy of spoken language as the original form of human communication (see e.g. Vachek (1939), but its status notwithstanding, spoken oral data is notoriously difficult to segment and further analyse. Researchers examining written texts can largely avail themselves of the safety of punctuation, and can thus delimit sentence boundaries with high precision and accuracy (see e.g. Lu 2010). Segmentation of oral data is, on the contrary, far less tangible, and apart from subordinating and coordinating conjunctions and other linking devices, researchers are left with general syntactic properties of the language and prosodic features. Several other criteria must be therefore considered to determine the boundaries of the production units. (Foster et al. 2000).

Despite the years of research and countless studies devoted to language complexity analysis, there has been an ongoing debate on the validity and reliability of both units of analysis and complexity measures, where none of them has gone uncriticised. As this chapter shows, research on the analysis of syntactic complexity of spoken L2 is characterized by a lack of a globally applicable method and typology of units of segmentation. So far, neither of the issues discussed in this chapter have been efficiently resolved, and no measures have been comprehensive and widespread enough to become a universal standard and make syntactic complexity analysis a more approachable enterprise. Needless to say, with the breadth of the extant research, such goal appears almost unfeasible.

#### 2.2.1.1 Units of Measurement

As it was shown in the previous sections, the research on language complexity is generally problematic in terms of its opaqueness. Issues mostly arise regarding the lack of definitions of the key concepts, and insufficient links between the theory, operationalization, and interpretation of the data. Such opaqueness is unfortunately also present in the area of data segmentation, where vague or missing definitions of units are not infrequent. Definitions are, beside this, often left without corresponding exemplification (Foster et al. 2000). This chapter provides an overview of the most influential units employed in spoken language research and shows the rigours that must be confronted when analysing any spoken text.

Until the end of the twentieth century, various units of segmentation came into existence. Spoken language analysis, although still an arguably minor research field, worked with units that were vague and did not encompass distinctive features of human speech. Other units (such as the T-unit) were used to analyse primarily written texts but were modified to suit spoken data. In 2000, Foster et al. published a breakthrough research survey, based on which they designed a robust unit for spoken language exclusively. The AS-unit (more in section 2.2.1.1.3) was created with the aim to be applicable to languages with syntax similar to that of English, which would ensure more reliable comparisons within the research field. They drew their findings from 87 studies to discover that not even half of them operated with no or a very vague definition, which was either not exemplified at all, or was accompanied by rather prototypical examples. The lack of detail veiled a range of problematic and borderline cases that are inevitable in spoken language analysis, and thus require systematic treatment.

Foster et al. discovered twelve units in use. The surprisingly high number has its roots in a conflicting theoretical background, at the core of which stands the dichotomy of spoken and written language. There were two opposing strands as to what units are appropriate for

segmentation of spontaneous spoken language. The first strand proposed identical syntactic units of analysis for both spoken and written language, and identified the sentence as the base unit (Miller & Weinert 2009). The contrasting view, held among others by Halliday (1989) and Miller & Weinert (2009) considered the sentence as inappropriate for spoken language analysis, assuming that spoken language is produced in building blocks of clause complexes. Miller & Weinert (ibid.) add that sentence is a "low-level discourse unit of spoken language" (p. 28), and that only clauses and phrases are shared between spoken and written language[2]. They argue that much of spoken production does not exhibit any traces of a syntactic structure, where the clause complexes could be analysed into respective clauses and phrases, and it is only sporadically that the complexes are syntactically linked (ibid.).

Designing a unit that allows for the specifics of speech is a tedious process. Spoken language can be highly elliptical and fragmentary; it is full of repetitions, self-corrections, false starts, and fillers (Foster et al. 2000, Miller & Weinert 2009). Apart from its distinct syntactic properties, speech is a direct product of micro- and macro- planning processes. It is now a prevalent assertion that microplanning activities produce shorter segments of speech in a length of a clause or sentence, while macroplanning activities are responsible for stretches spanning across the length of multiple sentences (Foster et al. 2000). The unit boundaries should as well respect intonation curves and pauses, but we must bear in mind that especially with dysfluent L2 speakers in question, intonation patterns and pauses might not be the most valuable criteria to delimit the unit boundaries (ibid.). Last but not least, memorized stretches of text such as idioms, formulaic expressions, and collocations have been given separate attention in research. They have an impact on fluency, as they can be retrieved quite effortlessly from the memory storage (Skehan 1998). Foster et al. (2000) mention them in connection with elaboration of

[2] Sentences have been largely abandoned in research due to their problematic nature for both written and spoken data (Foster et al. 2000).

syntax. It is debatable whether memorized stretches of text should be included in syntactic analysis, since their production requires no parsing.

The units reviewed by Foster et al. (2000) reflected the nature of spoken language in various amount of detail, but mostly to a limited extent. The authors considered units of both written and spoken language, which were divided into three groups based on their key constituting properties. These were semantic, intonational, and syntactic units (pp. 358 – 362). The most influential units are further elaborated on in the following sections.

#### 2.2.1.1.1 T-unit

The survey of Foster et al. (2010) proved that the *T-unit* had been the most widely used unit for both spoken and written data. Hunt (1965) designed the unit with the intention to avoid the sentence as the unit of analysis and consequently occurrences of disputable cases where delimiting the unit boundaries depends on subjective interpretation. Hunt (ibid.) identifies a T-unit as the "shortest grammatically allowable sentences into which (writing can be split) or minimally terminable unit" (p. 20). Up to 1970, Hunt produced four working definitions of the term:

- "one main clause with all subordinate clauses attached to it" (1965, p. 20)
- "one main clause plus whatever subordinate clauses happen to be attached to or embedded within it" (1966, p. 735)
- "the shortest units into which a piece of discourse can be cut without leaving any sentence fragments as residues" (1970, p. 185)
- "a main clause plus all subordinate clauses and non-clausal structures that are attached to or embedded in it" (1970, p. 4)[3]

[3] The definitions were cited in Foster et al. (2000, p. 360).

The observation of Foster et al. (2000) that researchers mostly use one of Hunt's definition is still valid nowadays, with the exception that it has ceased to be used in spoken L2 research, but remains generally accepted as the most satisfactory unit for analysis of written texts (Lu 2010). The research is, however, inconsistent as regards what is understood as a dependent clause. In an influential paper by Lu (2010), only finite clauses are considered as dependent clauses. Many researchers follow this trend also owing to the fact that this definition of a dependent clause was extended to Lu's (2010) and Kyle's (2016) automated analysers of written syntactic complexity, which are often preferred to manual analysis due to their high accuracy (more on these in section 2.3).

T-units are not particularly suitable for spoken data. As Foster et al. (2000) remark, only some of the working definitions address "sentence fragments" and "non-clausal structures", which means that they can be included or excluded based on the researchers' preferences. For written data, Schneider & Connor (1990) responded to this problem by adding punctuation criteria, allowing non-clausal structures and sentence fragments to be counted as a T-unit once they are punctuated as a sentence. For oral data, Young & Milanovic (1992) modified the definition to "an independent clause and any associated dependent clauses" (p. 409)[4], which on the one hand compensates for missing punctuation criteria, but generates even more questions as to, for example, the treatment of sentence fragments that are perfectly meaningful as an utterance within a dialogue, but constitute of one or two words. the T-unit has been consequently abandoned for oral data and replaced by more convenient units.

[4] The definitions were cited in Foster et al. (2000, p. 361).

#### 2.2.1.1.2 C-unit

The C-unit (or communication unit) is classified as a both syntactic and semantic unit by Foster et al. (2000). The reason behind the distinction is again a number of working definitions, which showed preference for different aspects of spoken language. The C-unit was designed by Loban (1966) as an offshoot of the T-unit, i.e. mostly a syntax-based unit, which also allows work with highly elliptical oral data. Loban (1966) describes the C-unit as following:

- 'grammatical independent predication(s) or … answers to question which lack only the repetition of the question elements to satisfy the criterion of independent predication. … 'Yes' can be admitted as a whole unit of communication when it is an answer to a question such as 'Have you ever been sick?' (pp. 5–6)[5]

Pica et al. (1989) developed the semantic dimension of the C-unit, defining it as 'utterances, for example, words, phrases, and sentences, grammatical and ungrammatical, which provide referential or pragmatic meaning' (p. 72)[6]. Bearing in mind that research on L2 production is essentially quantitative, this rather vague definition does not provide much insight as to how one should proceed in actual segmentation and how the 'pragmatic meaning' is defined and delimited.

Several other definitions were adopted up to 2000, some of them allowing the unit to include run-on sentence nodes, wasting the potential to work with oral data, while retaining clear syntactic boundaries. The major reservation Foster et al. (2000) express towards the C-unit definitions based on Loban (1966) is that they appear to exclude 'elliptical constructions which arise *within* a speaker's turn'(p. 316).

[5] The definition was cited in Foster et al. (2000, p. 361).
[6] The definition was cited in Foster et al. (2000, p. 361).

In the 20 years after Foster et al. (2000), C-unit has become a major unit in speech-language pathology research, appearing mainly in studies focusing on speech and language disorders, aphasia, and discourse level impairments (Bajaj 2007, Mäkinen et al. 2014, Bryant, Ferguson & Spencer 2016, Nippold et al. 2017 to name a few). Needless to say, coding and segmentation description has gained significant uniformity over the years, with software available for systematic analysis of transcripts and consistent coding and segmentation criteria (for more see section 2.2.1.1.5). The use of the unit is, however, not restricted to speech pathology research alone, it has appeared in a number of recent CAF-based L2 studies (see e.g. Lahmann et al. 2015, 2019, Bryfonski & Sanz 2018), although not as frequently as the AS-unit.

#### 2.2.1.1.3 AS-unit: The Final Answer?

Finally, Foster et al. (2000) identify their own unit, commonly referred to as the *AS-unit* (or analysis-of-speech-unit), which was designed exclusively for spoken language. The unit has been widely used since and has become the standard reference unit in spoken L2 CAF research along the T-unit for written language. Foster et al. (ibid.) define it as following:

- 'a single speaker's utterance consisting of an independent clause, or a sub-clausal unit, together with any subordinate clause(s) associated with either' (p. 365)[7]

The AS-unit is mainly syntactic and thus follows the tradition established by Hunt (1965) and the T-unit but is backed with extensive research in spoken L1 and L2 production. The psycholinguistic element of planning units was also considered – studies on the effect of pauses in L1 speech had shown that syntactic unit boundaries coincide with planning unit boundaries (ibid.). Moreover, the supra-clausal nature of the unit allows us to analyse more complex syntactic relations in our data. In their description of segmentation criteria, Foster et

[7] As in the original, the AS-unit boundaries are marked by an upright slash |, and the clause boundaries are marked by a double colon ::.

al. (ibid.) address and exemplify all the issues to which they had not found any solution in their survey, and reflect the specifics of spoken language in the instruction.

The AS-unit differs from other syntactic units in its treatment of sub-clausal units. A subclausal predication is considered an independent clause once it is either an elliptical utterance whose elided elements are fully recoverable from the context, or belongs to a class of *nonsentences* or *irregular sentences*, as it is defined in Quirk et al. (1985, pp. 838–853)[8]. A dependent clause can consist of both finite and non-finite predication once it contains a minimum of one additional element (usually an object or and adverbial).

In peculiar cases where it is especially difficult to delimit the borders of the unit, Foster et al. (ibid.) establish clear intonation and pause criteria. Such borderline clauses must be uttered within the same tone unit and cannot be preceded by a pause longer than 0.5 seconds for them to be included within the same AS-unit as the superordinate clause. It is usually optional subordinate adverbial clauses in final position and satellite noun phrases that fall within this category.

(1) | it's just a matter of passing (exams) and especially the basic education (0.5) | **they** have to pass automatically from one grade to another | (satellite NP, 2 AS-units) (p. 364)

(2) | specially for reading scientific papers | because er all the papers that er arrived to the library in Chile are English paper | (optional adverbial clause, 2 AS-units) (p. 368)

False starts, repetitions, and self-corrections are marked by curly brackets. It is not clear, however, why filled pauses are not given any attention in the coding instruction, considering that they would need to be systematically excluded from syntactic analysis. Foster et al. (ibid.) also establish segmentation criteria for interrupted speech and scaffolding in

[8] E.g. "yes", "thank you", "by all means". (Foster et al. 2000)

dialogic data, where a speaker is credited the same amount of AS-units as if they were not interrupted or corrected. The treatment of fragmentary units in highly interactive data is proposed in a three-level system of application, which allows the researchers to systematically exclude various amounts of texts depending on how fragmentary and elliptical their data is. Foster et al. (2000) provide several examples of coded dialogues, Ellis and Barkhuizen (2005) warn that the instruction may be sufficient for monologues, which are not as difficult to segment, but in dialogic data there is still room for subjective interpretation.

In spite of the wealth of examples and issues addressed, Foster et al. (2000) admit that there are important decisions that the researchers will have to make based on the character of their research and the nature of their data. Still, several problems have remained unaddressed and deserve a more systematic approach. A clear definition of a word is probably the most challenging task. For the sake of automated analysis, research most commonly works with a word as a string of phonemes that are meaningful when uttered in isolation, or a string of graphemes bound by spaces. Whilst it might be extremely difficult to establish more meaningful criteria, researchers must be aware of the risks of this definition and its likely consequences for syntactic complexity scores. A speaker referring to a film or a book title consisting of several orthographic words will produce a longer noun phrase or a clause, which translates as a higher complexity score length-based measures (see more in section 2.2.1.3). A solution that lends itself is the substitution of proper nouns by a uniform symbol, or including a series of fine-grained indices in the analysis, which would compensate for the inconsistencies in the word count. Proper nouns are but a fraction of more troubling phenomena that would need to be tackled as well, such as fixed phrases, idiomatic expressions, compound nouns, non-existent words, coinages, and many more. Another issue is the lack of attention to discourse markers of various length and their exclusion from the text, which will be dealt with in greater detail in the following section.

#### 2.2.1.1.4 Data Exclusion

The question of which data to exclude from the analysis is crucial for any kind of study that has to do with L2 production. The decision of which text to prune is essentially based on the nature of the data and the CAF component observed. A study measuring fluency in highly fragmentary dialogues will require a radically different treatment than a syntactic complexity analysis of monologues. Pruned and unpruned words defined already by Lennon (1990) in connection with fluency in L2: "pruned words exclude self-corrected words, repetitions (…) and words that may be classified as "asides", including comments on the narrating task itself and words addressed to the interviewer" (p. 405). Foster et al. (2000), criticising the general lack of principled approach in previous research, proposed a more complex system of data exclusion. It consists of three levels of application, where

- **Level 1** includes everything except untranscribable data and is meant for "a full analysis of all the data" (p. 370).
- **Level 2** is suitable for highly interactional data and excludes one-word minor utterances and verbatim echo responses.
- **Level 3** is only advised to be used in special cases, where one needs to analyse longer stretches of text and complete units. All the items in level 2 are excluded together with verbatim responses and elliptical responses where the elided text is recoverable from the interlocutor's turn. Finally, short greetings and closures are also excluded from the analysis.

This approach is undoubtedly more comprehensive than Lennon's definition and allows greater freedom in adjusting the data for a more accurate output, while being able to make use of an established framework of reference. In contrast, neither of the levels take into account discourse markers and 'asides', as Lennon (1990) defines them. Discourse markers vary in

length and range from single words to clauses. They help the speaker manage flow of discourse and are perfectly normal in spoken language. Yet, when used too frequently, they develop into a bad habit that has nothing to do with the speaker's ability to produce syntactically blocks of language.

Discourse markers are syntax-independent and are not integrated in the clause structure, which makes them separate AS-units. A speaker that ends or begins each utterance by "you know", or "I mean" will consequently produce a high number of very short units. Similarly to clausal discourse markers, single-word expressions like "so" or "okay", when included in the text, will again skew the length-based scores. To add to this, there will be borderline cases where it is not clear whether the chunk of text in question is a discourse marker or an integrated element, and only a principled treatment will ensure replicability of the study. The issue of discourse markers and asides and their treatment in spoken language segmentation has not yet been addressed in theory. Ellis & Barkhuizen (2005) provide a systematic overview of AS-unit segmentation rules, but problematic and borderline cases are only mentioned with respect to fluency analysis.

To discover how the research after Foster et al. (2000) responded to the appeal for systematic treatment of problematic data, twenty studies that use the AS-unit as the unit of analysis published from 2000 to 2019 were selected at random and reviewed in terms of their exemplification, coding and segmentation criteria, intercoder reliability, and their treatment of disputable cases (see Table 1). The analysis showed that vague definitions have been replaced by a new trend, which is a plain reference to Foster et al. (2000) for further information, mostly without exemplification. Only three studies (coloured green) met all or most of the criteria[9]. The study by Moser (2007) worked with the definition of AS-unit by Foster et al. (2000) and

[9] These studies referred to Foster et al. (2000) and used CHAT transcription format and coding criteria (MacWhinney 2020).

identified problematic data but proposed a completely new benchmark that has little in common with standard spoken language measures. Three studies (marked yellow) met only some of the criteria. The authors of two of them designed the AS-unit and understandably referred to their own account but did not elaborate on disputable cases. The rest of the studies only refer to AS-unit segmenting criteria and do not provide any coded text or information on data exclusion in supporting information.

| Study | AS-unit coding | Study | AS-unit coding |
|---|---|---|---|
| **Albert & Kormos (2004)** | Foster et al. (2000), gives an example of segmented text | **Magne et al. (2019)** | Foster et al. (2000) |
| **Fukuta, J. (2015)** | Foster et al. (2000) | **Michel, Kuiken, Vedder (2007)** | Foster et al. (2000) |
| **Ferrari, S. (2012)** | Foster et al. (2000) | **Moser (2007)** | Disputable cases in communicative data, proposes alternative benchmark |
| **Karayayla, T., & Schmid, M. S. (2018)** | Foster et al. (2000); CHAT; excludes formulaic expressions, gives examples of segmented text | **Pallotti (2014)** | Foster et al. (2000); comments on the treatment of subclausal units |
| **Kuiken & Vedder (2007)** | Foster et al. (2000) | **Skehan & Foster (2008)** | Foster et al. (2000) |
| **Lahmann et al. (2015)** | Forster et al.; CHAT; gives examples of pruned text; excludes formulaic expressions; gives examples of segmented text | **Santos (2018)** | Foster et al. (2000) High intercoder agreement but no report of disputable cases |
| **Lahmann et al. (2019)** | Foster et al. (2000) + CHAT criteria | **Takamasa et al. (2018)** | Foster et al. (2000) |
| **Lambert & Engler (2004)** | Foster et al. (2000) | **Tavakoli & Foster (2008)** | Foster et al. (2000) |
| **Lázaro-Ibarrola & Hidalgo (2017)** | Foster et al. (2000) | **Tavakoli & Skehan (2005)** | Foster et al. (2000); high intercoder agreement but no report of disputable cases |
| **Levkina & Gilabert (2012)** | Foster et al. (2000) | **Vercellotti (2018)** | Foster et al. (2000) |

*Table 1 Overview of studies with AS-unit as the unit of analysis*

#### 2.2.1.1.5 Other Coding and Segmentation Reference Systems

If a study requires a coding system, it is vital to specify details of coding method and segmenting conventions to ensure its replicability. Some of the studies reviewed in the previous sections made use of established formats of transcription of spoken texts and their respective coding. Codes for Human Analysis Transcript format (CHAT; MacWhinney 2020) is one of such tools. The coding system of the CHILDES project (Child Language Data Exchange System) was designed to provide "a standardized format for producing computerized transcripts of face-to-face conversational interactions" (MacWhinney 2020, p. 16) and offers a detailed instruction on coding titles and compound nouns, retracing, repetition, tone direction, and segmenting utterances and tag questions. The choice of the unit of analysis is, however, at the researchers' discretion[10]. Systematic Analysis of Language Transcripts (SALT) software was built with a similar aim, i.e. standardized and systematic treatment of language samples, but focuses on examining speech in connection with learning disabilities, language impairments, and developmental disorders. Unlike the CHAT format, the transcription and coding conventions recommend segmenting utterances into C-units. The manual offers detailed segmentation rules, which are based on Loban's (1976) definition of the unit (more in section 2.2.1.1.2) and include instruction on segmenting coordinated and subordinated clauses, sentence fragments, elliptical responses, and question tags.

A credit should also be given to Lahmann et al. (2019) from the University of Groningen, who designed extensive guidelines for annotating spoken syntactic complexity of L2 English and L1 German, where the annotation of syntactic complexity is assigned a special tier, instead of operating on the main tier with the utterance. The authors created the guidelines as part of the project "L1 attrition and L2 acquisition: a perspective from Kindertransport

[10] For example Lahmann et al. (2015) use the CHAT format combined with AS-units.

survivors", which focuses on L1 attrition in German Jews who emigrated to English-speaking countries during or after the WW2. The instruction is generally applicable to any syntactic complexity analysis of spoken English and German. If not as a coding instruction, then as a battery of segmented oral data.

#### 2.2.1.1.6 Conclusion

We should be aware that flaws are easy to find but extremely difficult to resolve. Research on spoken L2 has grown immensely and keeps generating new studies, which makes any prospect of updating segmentation and data exclusion criteria almost unfeasible. While the AS-unit is not without its faults, it still remains the most satisfactory and stable unit of spoken language analysis within the field of SLA. Foster et al. (2000) called for a systematic approach to the problems that have not yet been scrutinized. These include establishing clear data exclusion policy with sufficient exemplification in supporting data and high inter-coder agreement. Ideally, the final step towards high replicability would be creating a typology of AS-units and a unified system of data exclusion, but as the review showed, it is now crucial that the researchers start sharing descriptions of their method or that some general standards are set.

### 2.2.1.2 Measuring Syntactic Complexity

Syntactic complexity is the most frequently analysed dimension of L2 complexity (De Clercq & Housen 2017). In the decades of the research, numerous measures have been designed with the aim to gauge the construct. In 2005, over 80 different complexity measures excisted (Ellis & Barkhuizen 2005). A 2012 review of Boulté & Housen (2012) discovered that a greater part of the measures was used only in a handful of the studies[11]. Besides, the lack of

[11] Boulté & Housen (2012) analysed a total of 40 studies published from 1993 to 2008.

unified theory behind L2 complexity and the confusing polysemy of the construct have given rise to a body of research where either only scarce measures have been employed, or where the measures failed to respond to a clear complexity construct. This reductionist approach towards L2 complexity, as Pallotti (2015) identifies it, and the lack of connection between the theory and the operationalization have been criticized in many volumes, as were the attempts to reverse the trend and deliver more valid alternatives (see e.g. Norris & Ortega 2009, Pallotti 2009, Boulté & Housen 2012, Housen et al. 2012, Michel 2017).

Construct validity (i.e. "the extent to which measures adequately represent their underlying behavioural and theoretical constructs" Boulté & Housen 2012, p. 35), reliability of the measures, and comparability within the research field are the critical areas of enhancement that have been targeted in the most influential critiques (e.g. Wolfe-Quintero 1998, Norris & Ortega 2009, Pallotti 2009, Housen & Boulté 2012, Pallotti 2015). As Chapter 2.1 demonstrates, complexity studies have a tendency to work with vague, ambiguous definitions, which itself has an adverse impact on each of the areas. Boulté & Housen (2012) also discovered that researchers frequently measure the same sub-constructs of complexity several times using redundant metrics, instead of combining crude measures with finer-grained indices in order to identify the source of complexification. Norris & Ortega (2009) identified greatest redundancy in subordination metrics, which all have clauses in nominator and thus only measure subordination on different syntactic levels. To produce valid and comparable results, Norris & Ortega (ibid.) call for establishing "interpretation-centred warrants for what measures purportedly are measuring" (p. 570).

More recently, an issue has been raised of the validity of measures as markers of complexity development in L2 (Boulté & Housen 2012). The research was initially driven by the assumptions that any kind of complexity score will increase with language proficiency and that more complex generally stands for better language (Pallotti 2009). These were, however,

frequently challenged. Language development might not be linear and some structures can appear and disappear again from the learner's repertoire as their language competence develops (Vercellotti 2018). In addition, long utterances with excessive embeddings force the speaker and the listener to store multiple dependencies at once and thus obstruct fluent communication (Schiffrin 2014, Vercellotti 2018). Empirical studies on syntactic complexity development in L2 speech have produced mixed results, but mostly agreed that subordination increases with proficiency development (De Clerq & Housen 2017). Nonetheless, Vercellotti (2018) warns that data on L2 syntactic development have mostly been collected in cross-sectional studies, where "means can be skewed by a subset of the participants, and different subsets might drive 'significant' differences at different time points" (ibid., 5).

#### 2.2.1.3 Overview of Measures

The measures obtained in the surveys of complexity studies by Norris & Ortega (2009) and Boulté & Housen (2012) yield four basic groups. These comprise measurements of a) unit length, b) range of syntactic structures, c) structural complexity (sophistication), and d) the amount of coordination, subordination, and embedding, and are mostly calculated as frequencies, ratios, and formulas.

| Overall complexity | Sentential complexity | Subsentential (clausal & phrasal) complexity | Frequency-based complexity (sophistication) |
|---|---|---|---|
| 1. Mean length of unit (MLU)<br>2. Mean length of turn (MLT) | 1. Coordinated clauses/clauses<br>2. Clauses/unit<br>3. Dependent clauses/clause<br>4. Subordinate clauses /unit<br>5. Number of subordinate clauses<br>6. Subordinate clauses/dependent clauses<br>7. Relative clauses / unit | 1. Mean length of clause (MLC)<br>2. S-nodes/clause | 1. Frequency of passive forms<br>2. Frequency of conjoined clauses<br>3. Frequency of Wh-clauses<br>4. Frequency of imperatives<br>5. Frequency of conditionals<br>6. Frequency of auxiliaries<br>7. Frequency of comparatives |

*Table 2 Overview of syntactic complexity measures by Boulté & Housen (2012, p. 30—31)*

The surveys agree that while length-based and subordination measures are commonplace in syntactic-complexity studies, frequency-based sophistication measures rarely appear to complement them. Crudeness of measures has faced considerable criticism both in written and spoken complexity research for causing opacity of the measurements. Boulté & Housen (2012) express further concern about length-based metrics and subordination measures, since they capture two distinct theoretical complexity constructs and are therefore ambiguous. Subordination measures capture syntactic complexity as well as linguistic *difficulty* and are given greater weight in L2 complexity research based on the assumption that subordination is cognitively more demanding (ibid.)[12].

The hybridity of length-based metrics lies in their ability to measure not only syntactic, but also morphological complexity, given that the metrics are understood as a number of morphemes per production unit. Length-based metrics (or productivity measures, as Foster et al. (2000) & Vercellotti 2018 call them) are arguably the most robust and certainly the most frequent measures of syntactic complexity in both spoken and written complexity research. The basic logic behind the validity of the length-based measures is that with increasing proficiency, mean length of utterance also increases, and that the longer the utterance, the more complex its syntax (Ellis 1996, Boulté & Housen 2012). Length-based metrics have been used extensively in L1 research already since Brown (1973), whose empirical study on children's L1 development supported the assumption. While L2 studies have quickly adopted the measures, views opposing the reliance on length-based measures emerged (e.g. Pallotti 2009). Moreover, researchers have been paying attention almost exclusively to supra-clausal complexity (i.e. complexity of sentences, T-units, AS-units, or utterances) (Norris & Ortega 2009, Boulté & Housen 2012), which obscure the source of complexification. Norris & Ortega (2009) therefore

[12] While it may be possible that perceived difficulty of subordination might increase the validity of subordination measures, the link between cognitive complexity and linguistic complexity has not yet been confirmed (Boulté & Housen 2012).

proposed mean length of clause (MLC) as a complementary measure of phrasal complexity[13]. MLC was, however, criticized by Boulté & Housen (2012), who argued that a phrase can be complexified in several ways. If MLC solely is used to measure phrasal complexity, a hypothetical speaker producing 3 adjectives in a row will have the same complexity score as a speaker who modified their NP with a prepositional construction.

Structural complexity, which captures the width of repertoire of syntactic structures produced by a speaker, is a complexity subconstruct that has been given somewhat limited attention. Although structural complexity has been repeatedly suggested as a legitimate dimension of syntactic complexity (Foster et al. 2000, Norris & Ortega 2009), structural metrics have not yet been scrutinized enough to become conventional in L2 complexity research. Frequency-based ratios showed above may be useful for both spoken and written text. For oral data solely, Vercellotti & Packer (2016) suggested a proportional analysis of clause types as a fine-grained alternative to the categorization of written utterances as simple, compound, complex, and compound-complex. In addition, Vercellotti (ibid.) designed a weighted structural complexity scale that captures sophistication and acquisition of production units[14].

#### 2.2.1.4 Measuring Spoken Syntactic Complexity

Spoken L2 complexity has received notably less attention in research (Kuiken et al. 2019), perhaps also owing to the fact that it is much more difficult to operationalize than its written counterpart. There remain significant relationships to be established, especially between spoken complexity and language proficiency, as well as spoken and written complexity. For the measures employed in spoken L2 complexity research, Norris & Ortega (2009) and Boulté & Housen (2012) show that studies tend to use the quantitative means that are available for written

[13] Norris & Ortega (2009) identified three basic sub-constructs of syntactic complexity: general complexity (measured by length-based supra-clausal metrics), complexity by subordination (measured by metrics with clause in denominator), and subclausal (phrasal) complexity (measured by MLC).
[14] The scale was adopted in this research and is thus described in more detail in the method.

language. While the utility of identical holistic measures has been acknowledged in the spoken context, the necessity of fine-grained indices is still persistent (Vercellotti 2018, Lahmann et al. 2019). Also, the current approach to defining and operationalizing spoken complexity is not effectively applicable to very advanced speakers (Lahmann, et al. 2015).

The trajectory of L2 complexity development outlined in Norris & Ortega (2009) (see Chapter 2.2) indicates higher phrasal complexity in very advanced speakers[15]. This claim is, however, only valid for written language and more specifically for academic writing, where complex noun phrases are a key constituting feature. Clausal subordination is, as Biber et al. (2011) demonstrated, typical of daily informal conversation. Trebits (2014), however, asserts that speakers prefer very simple utterances due to high cognitive demands of spoken language production. A handful of studies pursuing the comparison of written and spoken syntactic complexity have shown that spoken language tends to be less complex (Kuiken et al. 2019), but even this research domain has largely generated inconclusive results, and the nature of the relationship between spoken and written syntactic complexity is still to be discovered (ibid.).

#### 2.2.1.4.1 Conclusion

Despite the rapidly growing volume of spoken syntactic complexity research, studies are still too scarce to yield any significant conclusions. Syntactic complexity of spoken language develops with increasing language proficiency, and it is now clear that the development is far from linear. It has not been proved that higher spoken complexity is directly related to higher language proficiency – the nature of spoken language poses a significant quantity of constrains for such proportionality to be possible. At this point, when an abundance of cruder measures and fine-grained indices is available, it is crucial to shift the focus to the

[15] A cross-sectional study (Lambert & Nakamura 2018) comparing syntactic complexity of 36 oral narratives performed by Japanese learners of English at intermediate and advanced level showed that structures that vary with proficiency levels are coordination, and nominal, adverbial, and relative subordination. The same variation was also found in L1 participants at different age levels.

areas of spoken language complexity that have not yet been sufficiently explored: namely the trajectory of spoken complexity development across all levels of proficiency, structural complexity, and spoken-specific measures reflecting distinctive features of speech.

## 2.3 Automatic Analysis of L2 Syntactic Complexity

Many issues that have been discussed with respect to operationalizing syntactic complexity emerged due to the laboriousness of manual computation, since there were no tools in existence that would automatize the metrics used in current research (Lu 2010). Until 2010, the only toolkit that included some syntactic complexity measures was Coh-metrix, developed by McNamara et al. (2004) to assess textual cohesion. With L2 syntactic complexity development in mind, Lu (2010) presented L2 Syntactic Complexity Analyzer (L2SCA), which offers automatic calculation of 14 of the most robust currently-used syntactic complexity indices, and thus means an immense contribution to L2 syntactic complexity research. L2SCA works with T-unit as well as sentence as the multiclausal units, and includes nine measures on the clausal or sentence level and four indices of phrasal complexity. The tool was developed on the corpus of written texts produced by advanced-level college students, and is intended to be used to analyse higher complexity levels (ibid.). A thorough analysis by Kyle (2016) showed very high accuracy of the analyser and highlighted its wide availability, the possibility of batch processing (which allows analysis of a large number of texts), and its portability. Kyle (ibid.), however, criticizes its "relatively sparse coverage" (p. 50) of measures. His Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASC) was designed for a broad analysis of L2 student writing and apart from Lu's (2010) 14 SCA indices, it includes thirty-one fine-grained indices of clausal complexity and seventeen fine-grained indices of phrasal complexity. TAASC indices mainly calculate average number of occurrences of the particular structure. Unlike L2SCA, TAASC does not operate online, but allows batch processing and is freely available for research. Analysis of spoken syntactic complexity has not

been fully automated because of numerous obstacles placed by the nature of spoken language and theoretical clashes in data coding and exclusions. Automated tools are, if not impossible, a matter of distant future. None of the studies on spoken complexity has, however, explored the extent to which L2SCA and TAASSC can analyse oral data, once the texts are formatted to correspond to the units defined by the analysts.

# 3 Research Questions

The aim of the thesis is to analyse productive and structural complexity of monologic tasks of B2 and C1 students of English philology with Czech as their L1 and determine whether (and how) syntactic complexity varies at different proficiency levels.

Originally, the thesis included another research question, where the aim was to discover whether available automated tools (i.e. L2SCA and TAASC) may be used to measure syntactic complexity of oral texts. As both are sensitive to punctuation in delimiting the unit boundaries, texts that were already divided to AS-units were adapted to resemble the written format[16]. It was clear that the analysers could not be possibly accurate with dialogic data, as their working definition of the T-unit does not account for sub-clausal structures and *nonsentences* (Kyle 2016). The thesis operated with the initial assumption that extensively pruned monologic data would have similar properties as written texts. However, the analysers are not entirely compatible with AS-unit coding criteria (Foster et al. 2000), in that their definition of clause does not include non-finite clauses. Three productive complexity measures (MLC, MLU, and C/As) were calculated manually and using L2SCA to measure accuracy of the analyser. The analysis, as expected, showed that manual and automated calculations were not highly

[16] Imposition of artificial punctuation criteria onto oral data may visually approximate written texts, but would also mean a critical interference into the meaning of the text. Therefore a compromise was opted for, where the AS-unit boundaries were marked by full stops instead of upward slashes and clause boundaries were removed altogether.

correlated (MLU r=.76; MLC, r =.39; C/As r = .69), and no further measures were computed using the tools. With respect to the impossibility to measure spoken syntactic complexity using L2SCA (Lu 2010) and TAASC (Kyle 2016), the research questions are formulated as follows:

RQ1: Is there a difference in productive syntactic complexity in informal speech of C1 and B2 advanced learners of English with Czech as their L1?

RQ2: Is there a difference in structural syntactic complexity in informal speech of C1 and B2 advanced learners of English with Czech as their L1?

# II. Empirical Part

# 4 The Dataset: LINDSEI_CZ

The data for the thesis derives from LINDSEI_CZ (Gráf 2017), a corpus of spontaneous spoken English compiled by PhDr. Tomáš Gráf PhD. as part of the LINDSEI project, a large-scale international corpus of advanced spoken learner English, which was launched in 1995 under the aegis of Centre for English Corpus linguistics at Université Catholique de Louvain. LINDSEI emerged as a spoken counterpart to the International Corpus of Learner English (ICLE), and contains oral data from advanced learners of English with various L1 backgrounds. To date, eleven subcorpora have been completed, and more are to be published (Cvrček & Richterová 2017).

The Czech component of LINDSEI (henceforth referred to as LINDSEI_CZ) contains 50 recordings of advanced speakers with Czech as their L1, which were compiled to conform with the LINDSEI 3-task format. In Task 1, speakers talk on a chosen topic; Task 2 constitutes an informal dialogue, and in Task 3 speakers reconstruct a story from four pictures presented to them without any time for preparation. The total duration of all three tasks is approximately 15 minutes (ibid.).

The subsample of LINDSEI_CZ analysed in this thesis consists of 20 monologic tasks (i.e. Task 1) of B2 (n=10) and C1 (n=10) speakers of English with Czech as their L1. All speakers were undergraduate students of English Philology in their final years of study. Participants were not administered any language proficiency test, as the recordings were subject to subsequent holistic ratings, which were made available together with the transcripts for the purposes of this study. The language proficiency levels were assigned to individual speakers through a series of proficiency descriptors that conform with the Common European Framework of Reference (Council of Europe 2001). The sample of 20 texts was selected based

on the length of the task, where the aim was to obtain continuous monologues that included only minimal amount of the interviewer's turns[17]. In Task 1, students were asked to choose one of three topics and talk about it for approximately 3-5 minutes. They were specifically instructed not to take notes, as that would obstruct the spontaneity of their speech (Gráf 2015). The instruction to the topics to prompt the elicitation was as following:

> *Topic 1: An experience you have had which has taught you an important lesson. You should describe the experience and say what you have learnt from it.*
>
> *Topic 2: A country you have visited which has impressed you. Describe your visit and say why you found the country particularly impressive.*
>
> *Topic 3: A film/play you've seen which you thought was particularly good/bad. Describe the film/play and say why you thought it was good/bad.*
>
> (Gilquin De Cock,& Granger 2010)

| Speaker | B2_1 | B2_2 | B2_3 | B2_4 | B2_5 | B2_6 | B2_7 | B2_8 | B2_9 | B2_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Tokens produced | 1040 | 297 | 450 | 559 | 671 | 626 | 495 | 707 | 819 | 583 |
| Speaker | C1_1 | C1_2 | C1_3 | C1_4 | C1_5 | C1_6 | C1_7 | C1_8 | C1_9 | C1_10 |
| Tokens produced | 753 | 566 | 535 | 1090 | 609 | 1006 | 428 | 511 | 596 | 744 |

*Table 3 Number of tokens produced by each speaker in Task 1*

The recordings were transcribed according to the LINDSEI system of transcription[18]. The transcription format is orthographic, contains no punctuation, and marks empty pauses in a three-tier system, which identifies pauses as "blank on the tape" in the length from below 1 second (a "short pause"), to a pause longer than 3 seconds (a "long pause"). "Medium pause"

[17] As Gráf (2015) reports, some participants admitted high anxiety levels, and thus needed more initiative from the interviewers even in the monologic tasks.
[18] The guidelines are available at https://uclouvain.be/en/research-institutes/ilc/cecl/transcription-guidelines.html.

is 1-3 seconds long. The transcription includes further information on filled pauses, backchanneling, truncated words, and voice quality.

# 5 Design & Methodology

## 5.1 Complexity Measures

Following the multi-dimensional approach of Norris & Ortega (2009), length-based and subordination metrics were employed to obtain a general picture of production complexity in individual speakers. Mean length of AS-unit, clauses / AS-unit, and mean length of clause were thus chosen for measuring overall, clausal, and phrasal complexity respectively. To prevent opacity of the results, fine-grained indices of structural complexity complement cruder measures. Firstly, following Biber et al. (2011) and Vercellotti & Packer (2016) the clauses were coded by clause types (main, coordinate, non-finite, adverbial, complement, relative). Latest studies on spoken complexity development and variation among proficiency levels (such as Vercellotti & Packer 2016, De Clercq & Housen 2017, Lambert & Nakamura 2018) show a tendency of coordinate and adverbial clauses to appear in the speaker's repertoire already in lower levels, while complement and relative clauses occur more frequently in the speech of advanced learners. Non-finite clauses tend to be coded as one group instead of assigning specific function roles (Vercellotti & Packer 2016), and therefore were not further classified in this analysis.

In addition to clause types named above, Vercellotti & Packer (2016) distinguish complement taking predicates (CTP) such as *think* or *guess*, and treat their dependent complement clause as the *matrix* clause of the structure. The theory behind the distinction lies in the claim that the complement clause in fact constitutes the semantic core of the utterance, while the function of the CTP is similar to a discourse marker (ibid.). This thesis, however, does not adopt this theory and uses the more traditional distinction, where such predicates constitute

the superordinate clause, and the matrix clause in the CTP theory constitutes its dependent complement clause. Complement-taking predicates were included in the coding process, but only for the sake of clarity – multi-clausal structures with *think* and *guess* are not as structurally complex as other complement clauses (De Clercq & Housen 2017), and the CTP score can therefore reveal potential overuse of the constructions.

Coding by clause types was complemented with a weighted scale of structural syntactic complexity designed by Vercellotti (2018), which is similarly based on the higher presumed complexity of relative and complement clauses.

| Score | Structure type |
|---|---|
| 0 | independent clause |
| 1 | coordinated verb phases; main clause + adverbial clause |
| 2 | main clause + non-finite clause |
| 3 | main clause + relative clause; main clause + complement clause |
| 4 | main clause + combination of 2 or more dependent clause types |

*Table 4 Weighted structural complexity scale (Vercellotti 2018)*

## 5.2 Data Processing

The transcripts were adjusted for the subsequent semi-automatic analysis of syntactic complexity. The changes made in the texts included redefining criteria for marking pauses in the texts, segmenting the texts into units of analysis (i.e. clauses and AS-units), and excluding data incompatible with syntactic complexity analysis of oral monologic texts.

### 5.2.1 AS-unit Coding

AS-unit by Foster et al. (2000) was chosen as the unit of analysis. The thesis adopts the definition of the unit and the subsequent segmenting criteria: "AS-unit is a single speaker's utterance consisting of an *independent clause*, or *subclausal unit*, together with any *subordinate*

*clause(s)* associated with either" (p. 365). In the transcripts, unit boundaries are marked with upward slashes "|" and clause boundaries within the unit are marked with double colons "::". False starts are included in curly brackets "{}". Constituting elements of AS-units were segmented according to the following rules:

*An independent clause* must include at least one finite verb.

1a | so she won |
1b | I'm not sure |
1c | and it's not about that the particular place or the economical situation |

As subordination is one of the most significant dimensions in syntactic complexity research, it is necessary to pay special attention to the definition of a subordinated clause. Here a *dependent clause* consists of a minimum of one finite or non-finite verb element and at least one more other constituting element (such as object, adverbial, or subject complement)[19]. The criterion of an additional element was established to prevent cases such as the gerund in "I love reading" from being coded as dependent clauses (ibid., 365).

2a | so she felt I don't know offended :: that we are in her house | (1 AS-unit, 2 clauses)
2b | it's my plan for the future :: to return to London | (1 AS-unit, 2 clauses)
2c | so the London Eye had to go back | (1 AS-unit, 1 clause)

This entails that inserted clauses such as "you know", "I don't know", "let's say", were not considered separate clauses but were counted as words within the clause where they were inserted. This measure ensures that a speaker who has a tendency to overuse such clauses will not be assigned lower clausal length-based complexity score. At the same time, a disadvantage of this treatment is that the same speaker will consequently have higher phrasal complexity

[19] This distinction is especially important with respect to automated analysis tools such as L2SCA (Lu 2010), whose definition of a clause includes only finite verbs.

score, given that MLC is employed as a measure of phrasal complexity (as suggested by Norris & Ortega 2009).

An AS-unit can be further constituted by *sub-clausal units*, which Foster et al. (2000) define either as elliptical phrasal utterances that can be recovered into a full sentence from the context, for example in a dialogical exchange, or as minor utterances identified by Quirk et al. (1985) as "nonsentences" or "irregular sentences". Subclausal units were, however, removed from the analysis due to the nature of the task (for more on data exclusion, see section 5.2.2).

3a | yeah |

3b | not really |

Coordination posed a specific problem in determining the unit boundaries. Coordinated clauses were split into two separate AS-units by default. In contrast, coordinated verb phrases could be retained in one AS-unit unless there was a clear drop or rise in intonation in the first of the coordinated phrases, or a pause longer than 0.5 seconds. This means that LINDSEI three-tier pause marking system had to be transformed to fit the AS-unit pause criteria. Foster et al. (2000) argue that the pause of 0.5 seconds is long enough to be reliably measured without special tools, and that clauses divided by such a pause and change in intonation are part of the following utterance.

4 | and then she stops :: and like grabs I don't know a ball of the snow | it's an imaginary one (0.5) | and eats it :: again imitating it |

Some speakers showed a tendency to end their utterances with coordinating conjunctions, which were then functionally similar to discourse markers. Such conjunctions were recognizable by a noticeable pause and dropping intonation, and were thus retained in the first of the units. Moreover, some of these were followed by another coordinating element.

5 | the movie was about seven hours long as well so (0.5) | but I really like it so (0.5) |

Dependent clauses can realize several functions. These are most often complement, relative, and adverbial. While initial and medial adverbial clauses are not particularly problematic (example 6a), dependency of optional adverbial clauses in final position is somewhat blurred (example 6b). For these cases, Foster et al. (ibid.) established a rule that an adverbial clause must be within the same tone unit as at least one of the preceding elements for it to stay within the same AS-unit.

6a | they were helpful :: even when we were not lost | (1 AS-unit)

6b | and from there I went to Toronto :: which was very beautiful | because obviously the city is huge | (2 AS-units)

False starts as defined in LINDSEI transcription manual were extended to include remarks similar to Lennon's (1990) asides, in that they are not part of the story told in the monologue, but comment on the student's inability to recollect a term in the target language. If these were uttered after an incomplete clause that was missing a constituting element, the incomplete clause was marked as a false start.

7 | {and we saw} I 'm not sure about the real name of this building |

The final criterion pertains to the definition of a word. This analysis uses the orthographic definition of the word (more in section 2.2.1.1.3). Although it is clear that this is by no means an ideal solution, it is not the purpose of this thesis to design a new coding system, and the orthographic understanding of a word is the most common in the research because of its convenience in automatic and semi-automatic analysis.

### 5.2.2 Data Exclusion

Data unfit for the analysis were excluded in several stages. Firstly, texts were adapted to exclude interviewer's turns, false starts, repetitions, filled pauses, backchannelling, and

information on voice quality, which were deleted after AS-units segmentation, since they could help delimiting unit boundaries in less clear cases. As this thesis aims to measure continuous stretches of texts and relatively complete units, it has adopted Level 3 of data exclusion by Foster et al. (2000) (for the description Level 3 exclusion, see section 2.2.1.1.4). The definition of verbatim echo responses was extended to include the openings of the speeches, where students read the instruction on the topic. In several instances, the interviewer clearly signals the beginning of the task. The section of the text preceding such signal was then excluded from the analysis:

8 A: hello <name of the interviewee> right what are you going to speak to us about
B: hi I've chosen the topic three I should probably read out the instruction right yeah
A: yes read out the topic and
B: I will be talking about a film or a play that I've seen which I thought was particularly good or bad and then describe it
A: great off we go[20]

The initial comments of the speaker in the excerpt above can be classified as asides in Lennon's (1990) terminology. These were, however, not excluded text-wide, but only when clearly preceding the monologic task (such as in example 8, cf. example 7). Excluding longer stretches of text within the task solely on semantic criteria would interfere with the nature of the AS-unit, which is primarily syntactic. AS-units that were missing obligatory elements were removed from the analysis as well. With respect to the monologic nature of the speeches, all dialogic exchanges were similarly excluded. De Clercq & Housen (2017) recommend this procedure in monologic tasks, since responses of the speaker in dialogic exchanges are elliptical to various extent.

9 B: nothing is open so we had a little problem with water but
A: on an island in the sea that's funny
B: yes it was

[20] "A" marks the interviewer's turns; the speaker's turns are marked by "B".

Finally, following excerpts illustrate the extent of data exclusion in the analysis. Example 10a shows the original proportion of the text, where the words in italics were excluded. 10b is the final version of the text adapted for the complexity analysis:

10a A: *hello <first name of interviewee> welcome so what have you decided to talk about*
B: *erm I've decided to talk about topic one*
A: *okay*
B: *em an experience that have taught me something*
A: *good*
B: and for me this experience *eh* would be *em mm the choosing of* the choice of university *after* after high school
A: *(mhm)*
B: because it's an important decision and I've made a wrong decision *but I think*
A: *do you think so*
B: *yeah*
A: *tell us about it*
B: I think I've learnt from it

10b | and for me this experience would be the choice of university after high school | because it's an important decision | and I've made a wrong decision | I think I've learnt from it |

# 6 Results

The following section presents the results of the analysis comparing syntactic complexity of B2 and C1 (20 samples in total) monologic tasks derived from LINDSEI_CZ. Productive complexity is operationalized as mean length of clause (i.e. mean number of words per clause; MLC), mean length of AS-unit (i.e. mean number of words per AS-unit; MLU), and the ratio of clauses / AS-units (C/AS). The combination of length-based and subordination measures (i.e. crude measures) follows the recommendation of Norris & Ortega (2009) and Pallotti (2015) to analyse syntactic complexity on phrasal, clausal, and supra-clausal level. Structural complexity is operationalized as clause-type / clause ratio (fine-grained indices) and the proportion of structures corresponding to five tiers of weighted complexity scale. Tables 5 and 6 provide the basic descriptive statistics detailing the absolute frequencies of all of the items measured.

*Table 5 Numbers of clauses, AS-units, clause types, and complexity scale structure produced by C1 speakers*

| Speaker | C1_1 | C1_2 | C1_3 | C1_4 | C1_5 | C1_6 | C1_7 | C1_8 | C1_9 | C1_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Clauses | 116 | 91 | 75 | 170 | 81 | 141 | 68 | 86 | 82 | 112 |
| AS-units | 79 | 52 | 50 | 115 | 36 | 94 | 36 | 44 | 49 | 68 |
| **Clause types** | | | | | | | | | | |
| Independent clauses | 57 | 24 | 29 | 78 | 12 | 60 | 16 | 21 | 32 | 40 |
| Adverbial clauses | 6 | 5 | 9 | 17 | 14 | 8 | 5 | 9 | 9 | 4 |
| Coordinate VPs | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 1 |
| Complement clauses | 12 | 16 | 4 | 20 | 14 | 21 | 9 | 11 | 11 | 13 |
| Relative clauses | 12 | 5 | 9 | 8 | 5 | 7 | 9 | 8 | 7 | 17 |
| CTP | 2 | 7 | 0 | 1 | 3 | 5 | 1 | 1 | 0 | 2 |
| Non-finite clauses | 3 | 13 | 2 | 7 | 9 | 9 | 8 | 10 | 5 | 5 |
| **Complexity scale** | | | | | | | | | | |
| Level 0 structures | 57 | 24 | 29 | 78 | 12 | 60 | 16 | 21 | 32 | 40 |
| Level 1 structures | 4 | 0 | 6 | 9 | 7 | 5 | 2 | 3 | 2 | 4 |
| Level 2 structures | 2 | 6 | 1 | 2 | 2 | 8 | 2 | 4 | 2 | 2 |
| Level 3 structures | 12 | 14 | 8 | 18 | 3 | 17 | 7 | 5 | 7 | 18 |
| Level 4 structures | 4 | 7 | 4 | 8 | 10 | 2 | 4 | 9 | 6 | 2 |

*Table 6 Numbers of clauses, AS-units, clause types, and complexity scale structure produced by B2 speakers*

| Speaker | B2_1 | B2_2 | B2_3 | B2_4 | B2_5 | B2_6 | B2_7 | B2_8 | B2_9 | B2_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Clauses | 166 | 40 | 65 | 78 | 95 | 106 | 71 | 108 | 139 | 89 |
| AS-units | 133 | 25 | 28 | 49 | 68 | 78 | 46 | 62 | 71 | 52 |
| **Clause types** | | | | | | | | | | |
| Independent clauses | 101 | 14 | 9 | 31 | 47 | 56 | 31 | 32 | 35 | 28 |
| Adverbial clauses | 7 | 3 | 11 | 8 | 7 | 4 | 4 | 7 | 18 | 5 |
| Coordinate VPs | 0 | 0 | 0 | 0 | 4 | 0 | 6 | 4 | 1 | 1 |
| Complement clauses | 10 | 0 | 13 | 10 | 4 | 12 | 3 | 13 | 25 | 13 |
| Relative clauses | 1 | 5 | 5 | 4 | 6 | 3 | 5 | 8 | 8 | 10 |
| CTP | 0 | 0 | 0 | 2 | 0 | 5 | 0 | 4 | 6 | 0 |
| Non-finite clauses | 2 | 8 | 2 | 5 | 4 | 8 | 8 | 15 | 13 | 7 |
| **Complexity scale** | | | | | | | | | | |
| Level 0 structures | 101 | 14 | 9 | 31 | 47 | 56 | 31 | 32 | 35 | 28 |
| Level 1 structures | 7 | 2 | 3 | 3 | 5 | 2 | 1 | 3 | 5 | 1 |
| Level 2 structures | 2 | 4 | 2 | 1 | 1 | 5 | 4 | 9 | 5 | 3 |
| Level 3 structures | 11 | 2 | 4 | 9 | 10 | 14 | 7 | 13 | 7 | 13 |
| Level 4 structures | 2 | 2 | 9 | 5 | 5 | 2 | 3 | 3 | 18 | 7 |

## 6.1 Productive Complexity – Results

Productive complexity scores were obtained following a semi-automatic analysis[21] and reflect the procedure of division into units of analysis and systematic data exclusion presented in section 5.2.1 and 5.2.2.

As Table 7 shows, the datasets are almost identical in each of the measures. The C1 speakers may have higher mean scores in all three metrics, but only with very small differences (spanning from the difference of .04 words in MLC, to .06 clauses per AS-unit, to .48 words in MLU). Mean length of clause with the maximum score difference of .13 words / clause between the median values is the measure where the uniformity of the data is the most noticeable. The subordination scores (clauses / AS-unit) are again slightly higher in C1 group, but generally do not exceed the threshold of 2 clauses per AS-unit. This indicates that a greater part of the AS-units produced (irrespective of proficiency level) are independent clauses.

| | | Min. | Max. | Mean | STD | Median |
|---|---|---|---|---|---|---|
| MLC | B2 | 5.89 | 7.43 | 6.67 | .53 | 6.73 |
| | C1 | 5.94 | 7.52 | 6.71 | .52 | 6.56 |
| MLU | B2 | 7.82 | 16.07 | 11.00 | 2.30 | 11.30 |
| | C1 | 9.48 | 16.92 | 11.48 | 2.10 | 10.91 |
| Clauses / AS-unit | B2 | 1.25 | 2.32 | 1.65 | .31 | 1.59 |
| | C1 | 1.47 | 2.25 | 1.71 | .26 | 1.66 |

*Table 7 Productive complexity – descriptive statistics*

[21] The number of words in each clause and AS-unit, and the number of clauses in each AS-unit were calculated using a computer script.

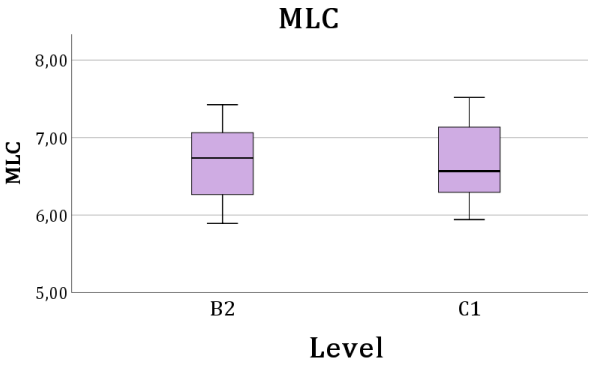



*Figure 1 Distribution of MLC scores*

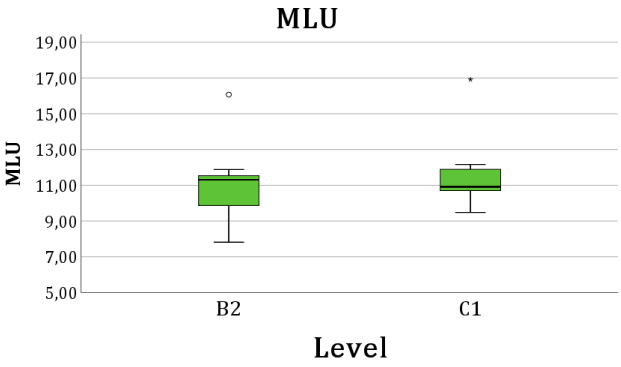



*Figure 2 Distribution of MLU scores*

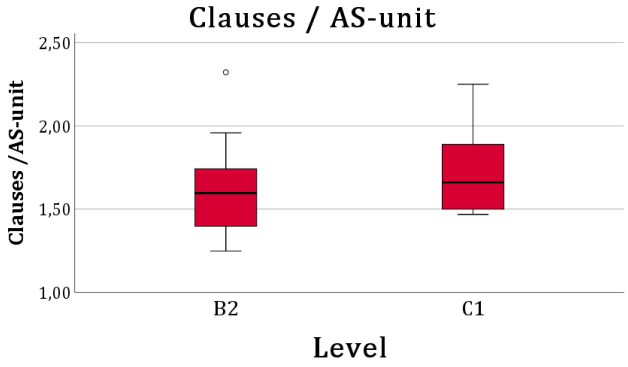



*Figure 3 Distribution of Clauses / AS-unit scores*

Figures 1–3 facilitate the comparison of the distribution of MLC, MLU and clauses / AS-unit scores in B2 and C1 speakers. Fig. 1 shows that despite the higher mean, minimum, and maximum scores of C1 group, higher median value of B2 MLC scores and the distribution of the lower 50% indicate that B2 speakers in fact tended to produce somewhat longer sentences than C1 speakers. Figure 2 comparing MLU scores revealed greater consistency of unit length in the C1 group. While both B2 and C1 maximum values reach above 16 words per AS-unit, units of such length were only produced by one speaker in each group. These speakers maintained consistently higher productive complexity scores (see speakers B2_5 and C1_3 in Appendix 1), but their structural complexity scores (see section 6.2, Appendix 2, and Appendix 3) were rather average.

| | MLU | MLC | Clauses / AS-unit |
|---|---|---|---|
| U | 47.00 | 46.00 | 43.00 |
| Z | -.23 | -.30 | -.52 |
| P | >.05 | >.05 | >.05 |

*Table 8 Mann–Whitney U Test scores comparing productive complexity scores for of and C1 speakers*

To compare MLU, MLC, and C/AS scores of B2 and C1 group, MannWhitney U Test was carried out.[22] The test showed no significant difference between the groups in any of the large-grained complexity measures.

## 6.2 Structural Complexity – Results

Structural complexity scores were similarly obtained following a semi-automatic analysis, where the number of occurrences per each clause type was counted in the coded texts using a computer script, and subsequently divided by the total number of clauses to normalize the data. Table 9 shows the proportion of each clause type, where

ADV = an adverbial clause
NONF = non-finite clause
REL = relative clause
COOR = coordinate predicates (compound verb phrase)
COMP = complement clause
INDEP = independent clause
CTP = complement-taking predicate

| | ADV | | NONF | | REL | | COOR | | COMP | | INDEP | | CTP | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B2 | C1 | B2 | C1 | B2 | C1 | B2 | C1 | B2 | C1 | B2 | C1 | B2 | C1 |
| Min. | .04 | .04 | .02 | .03 | .01 | .05 | .00 | .00 | .00 | .05 | .16 | .15 | .00 | .00 |
| Max. | .20 | .17 | .20 | .14 | .13 | .16 | .08 | .02 | .23 | .17 | .77 | .50 | .09 | .14 |
| Mean | .09 | .08 | .09 | .08 | .07 | .09 | .02 | .00 | .11 | .13 | .40 | .35 | .03 | .04 |
| STD | .05 | .04 | .05 | .04 | .03 | .06 | .03 | .01 | .07 | .03 | .16 | .11 | .04 | .04 |
| Median | .07 | .09 | .08 | .06 | .07 | .09 | .00 | .00 | .12 | .14 | .38 | .38 | .00 | .03 |

*Table 9 Proportion of individual clause types – descriptive statistics*

[22] Due to the small size of the sample, non-parametric statistical tests were performed to analyse both productive and structural complexity.

Most of the clause types are marked by higher dispersion of B2 scores. Independent clauses account for the largest part of the text irrespective of the speakers' level of proficiency, which accords with the relatively low subordination score (clauses / AS-unit; cf. Table 7).

The boxplots show higher dispersion of data in the B2 group, where the scores range from 16% to 77% of the clauses produced. Mean and median values indicate that lower 50% in both groups produced similar proportions of independent clauses. In the C1 group, independent clauses comprise up to 50% of the texts, but in most part do not exceed 40%.

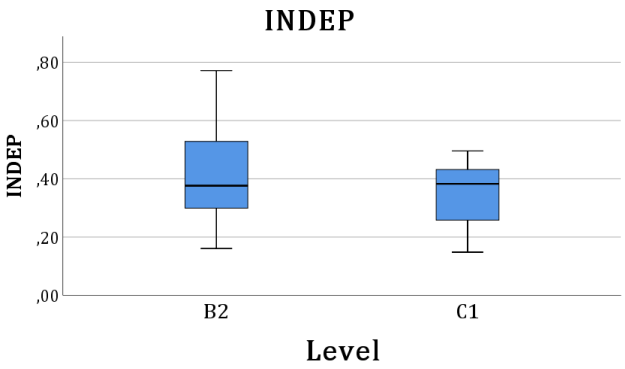



*Figure 4 Proportion of independent clauses in the texts*

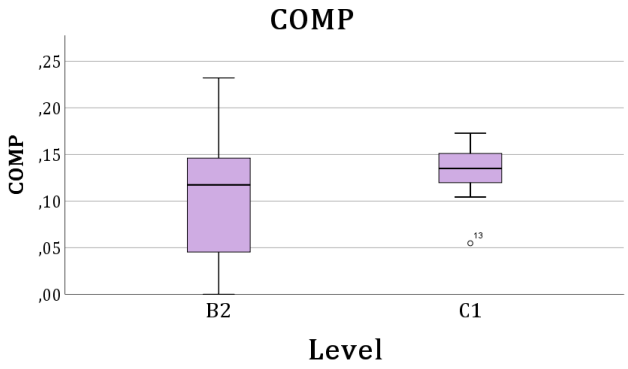



*Figure 5 Proportion of complement clauses in the texts*

Complement clauses were the second most frequent clause type used among the speakers and is again marked by a greater consistency of the data in the C1 group. B2 speakers mostly produced from 5% to 15% complement clauses, with an exception of speaker B2_3, whose complement clauses comprise 23% of the clauses in the monologue. To discover the extent to which these were introduced by complement-taking predicate (e.g. *guess, think*), the proportion of CTP within COMP was calculated for each speaker.

| Speaker | B2_1 | B2_2 | B2_3 | B2_4 | B2_5 | B2_6 | B2_7 | B2_8 | B2_9 | B2_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| COMP | .08 | .00 | .23 | .13 | .05 | .11 | .04 | .12 | .18 | .15 |
| CTP | .00 | .00 | .00 | .03 | .00 | .06 | .00 | .07 | .09 | .00 |
| Ratio | .00 | .00 | .00 | .23 | .00 | .56 | .00 | .55 | .47 | .00 |

*Table 10 Proportion of CTPs in structures with complement sentences: B2 speakers*

| Speaker | C1_1 | C1_2 | C1_3 | C1_4 | C1_5 | C1_6 | C1_7 | C1_8 | C1_9 | C1_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| COMP | .10 | .17 | .05 | .12 | .17 | .15 | .13 | .14 | .14 | .12 |
| CTP | .03 | .14 | .00 | .01 | .09 | .05 | .03 | .02 | .00 | .03 |
| Ratio | .24 | .80 | .00 | .07 | .51 | .36 | .24 | .18 | .00 | .25 |

*Table 11 Proportion of CTPs in structures with complement sentences: C1 speakers*

The calculations revealed that the B2 speaker with the largest proportion of complement clauses did not, in fact, introduce any of them with a complement-taking predicate. In contrast, CTPs were associated with approximately 50% of the complement clauses in three other speakers, whose COMP scores are similarly high. This indicates that B2_3 speaker produced generally more complex complement clauses. Six out of ten B2 speakers did not use any complement-taking predicates. In C1 group, eight out of ten speakers included CTPs in their COMP structures, with the proportion of CTP up to 80%.

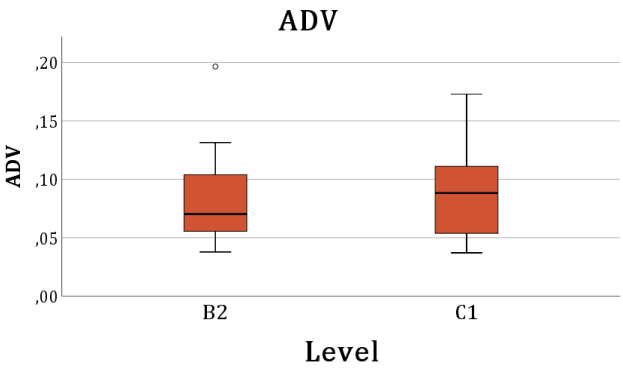



*Figure 6 Proportion of adverbial clauses in the texts*

Figure 6 shows the distribution of adverbial clauses in B2 and C1 speakers, where the majority of the speakers produced a very similar proportion of the clause type. An outlier in the B2 boxplot represents speaker B2_3, who scored very high also in complement clauses. A more thorough analysis revealed that 75% of adverbial clauses produced by the speaker were adverbial clauses of reason introduced by conjunction *because*. This indicates overuse of the structure, which is a frequent phenomenon in spoken texts (Foster et al. 2000). In contrast, speaker C1_5 with the

proportion of 17% used a wider repertoire of adverbial clauses, where only 21% of them were introduced by *because*.

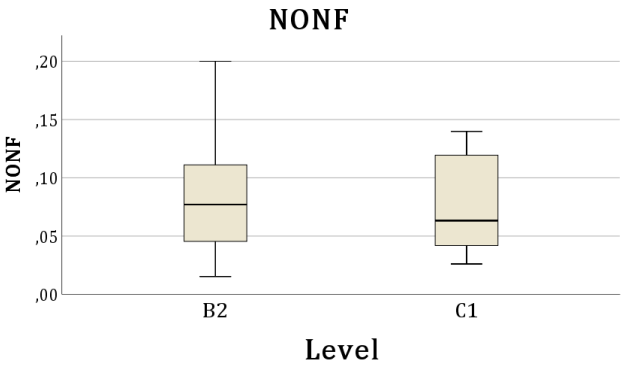



*Figure 7 Proportion of non-finite clauses in the texts*

Non-finite and relative clauses were among the least frequent subordinate clause types. Both types are marked by overall variation and slightly higher scores in the C1 group. The highest scores in the B2 group are represented by the same speaker (B2_2), whose NONF and REL scores are contrasted with no complement clauses and coordinated predicates. This might indicate that idiosyncrasy plays an important role in structural complexity.

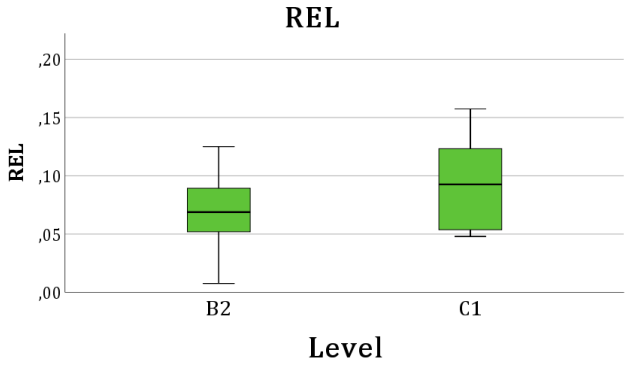



*Figure 8 Proportion of relative clauses in the texts*

Coordination scores were obtained solely from coordinated finite verb phrases (i.e. with elliptical subject). The decision to exclude coordinated clauses (with overt subjects) from the analysis was made to account for the AS-unit segmenting criteria,

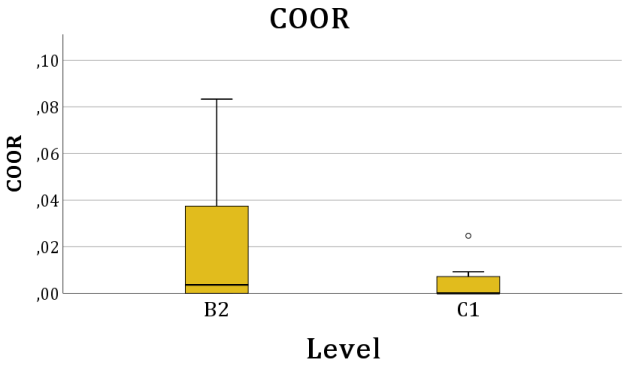



*Figure 9 Proportion of coordinated predicates in the texts*

which divide coordinated clauses into separate units by default and thus prevent measuring coordination in the cases where a clause introduced by a coordinating conjunction clearly marks the beginning of new tone unit. The same criterion was applied to coordinated verb phrases as well, which left only a marginal portion of such constructions for analysis. Moreover, speakers

frequently use coordinating conjunction as fillers rather than actual connecting devices (De Clercq & Housen 2017). Figure 9 points at a very marginal use of coordinated predicates (with median values at .00). It may appear that B2 speakers used complex verb phrases to a much greater extent than C1 speakers, but the percentage difference is still negligible.

| | ADV | NONF | REL | COMP | COOR | INDEP | CTP |
|---|---|---|---|---|---|---|---|
| Mann–Whitney U | 41.500 | 48.500 | 35.000 | 34.500 | 38.500 | 43.500 | 37.00 |
| Z | -.64 | -.11 | -1.13 | -1.17 | -.95 | -.49 | -1.02 |
| P | >.05 | >.05 | >.05 | >.05 | >.05 | >.05 | >.05 |

*Table 12 Mann–Whitney U Test scores for clause types*

Mann–Whitney U test was performed to compare individual clause-type scores in B2 and C1 speakers. Again, the test revealed no significant difference between the groups in any of the measures.

Scores for weighted structural complexity scale were obtained using a computer script and adapted in the same manner as in previous measures.

| | 0 | | 1 | | 2 | | 3 | | 4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Level | B2 | C1 | B2 | C1 | B2 | C1 | B2 | C1 | B2 | C1 |
| Min | .33 | .35 | .02 | .00 | .01 | .02 | .08 | .09 | .02 | .02 |
| Max | .82 | .72 | .11 | .21 | .17 | .12 | .25 | .27 | .33 | .29 |
| Mean | .60 | .58 | .06 | .08 | .08 | .06 | .16 | .18 | .11 | .11 |
| SD | .14 | .11 | .03 | .06 | .05 | .04 | .06 | .06 | .10 | .09 |
| Median | .61 | .61 | .06 | .06 | .07 | .05 | .15 | .10 | .08 | .03 |

*Table 13 Proportion of complexity-scale levels – descriptive statistics*

Table 13 shows that level 0 structures were by far the most common. As level 0 included only independent clauses, the distribution of the data corresponds with that of INDEP. Level 3 structures were the second most common and include the combined score of structures comprising of matrix + relative clause and matrix + complement clause. Level 4 structures follow (matrix clause + minimum of 2 clause types), which implies that speakers tended to

combine independent clauses with more advanced structures at the expense of the structures that are typical for lower-level students (Vercellotti 2018).

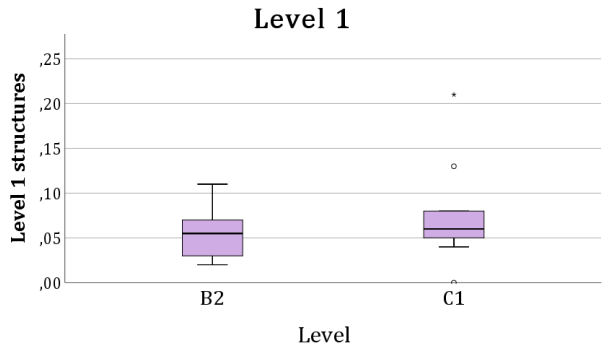



*Figure 10 Proportion of Level 1 structures in the texts*

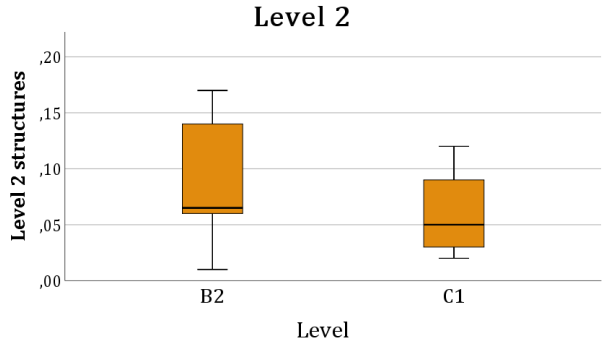



*Figure 11 Proportion of Level 2 structures in the texts*

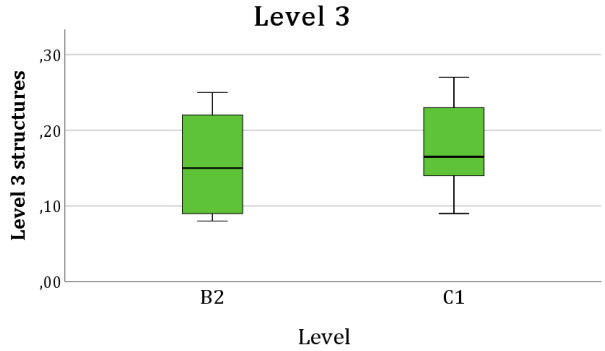



*Figure 12 Proportion of Level 3 structures in the texts*

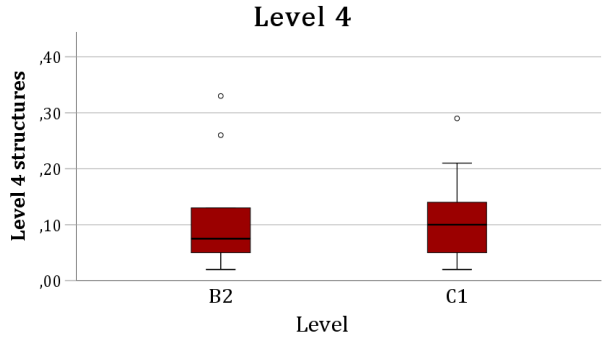



*Figure 13 Proportion of Level 4 structures in the texts*

Figures 10–13 illustrate the differences in the groups across the complexity scale levels. The distribution is very similar in nearly all levels, B2 speakers outscoring the C1 group only in L2 level structures. Considering the outliers in Level 1 and Level 4 boxplots, three out of ten speakers in the C1 group produced significantly higher or lower scores than the rest of the group in Level 1 structures, which illustrates variability of the data rather than a group trend. Figure 12 reveals that despite the higher overall score of C1 in Level 4 structures, two B2 speakers outperformed the rest of the group with 26% (B2_9) and 33% (B2_3) (for more see Appendix 3). Speaker B2_3 has been mentioned in connection with their extensive use of adverbial and complement clauses. The fact that the speaker's scores are above average in all subordinate clause types indicates that the highest syntactic complexity score in fact belongs to a lower-level speaker.

| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Mann–Whitney U | 43.50 | 41.00 | 36.50 | 40.00 | 46.50 |
| Z | -.50 | -.69 | -1.03 | -.77 | -.27 |
| p | >.05 | >.05 | >.05 | >.05 | >.05 |

*Table 14 Mann–Whitney U Test scores for Weighted complexity scale*

Mann–Whitney U test was carried out to compare scores across all complexity scale levels in B2 and C1 speakers. The analysis yielded no significant effect of language proficiency on structural complexity scores. To test the assumption that speakers irrespective of language proficiency level used a significantly higher number of advanced structures (level 3 and 4) than lower-level structures (level 1 and 2), Mann–Whitney U test was performed and revealed a significant difference between the levels ($U = 339$, $Z = 4.45$, $p < 0.01$). The analysis, however, did not prove that C1 speakers would use s significantly higher number of advanced structures (level 3 and 4) than B2 speakers ($U = 187$, $Z = .354$, $p > .05$).

Since the statistical tests revealed no significant difference between the samples in any of the structural complexity variables, Principal Component Analysis (PCA) was then performed to reduce data-set dimensionality.

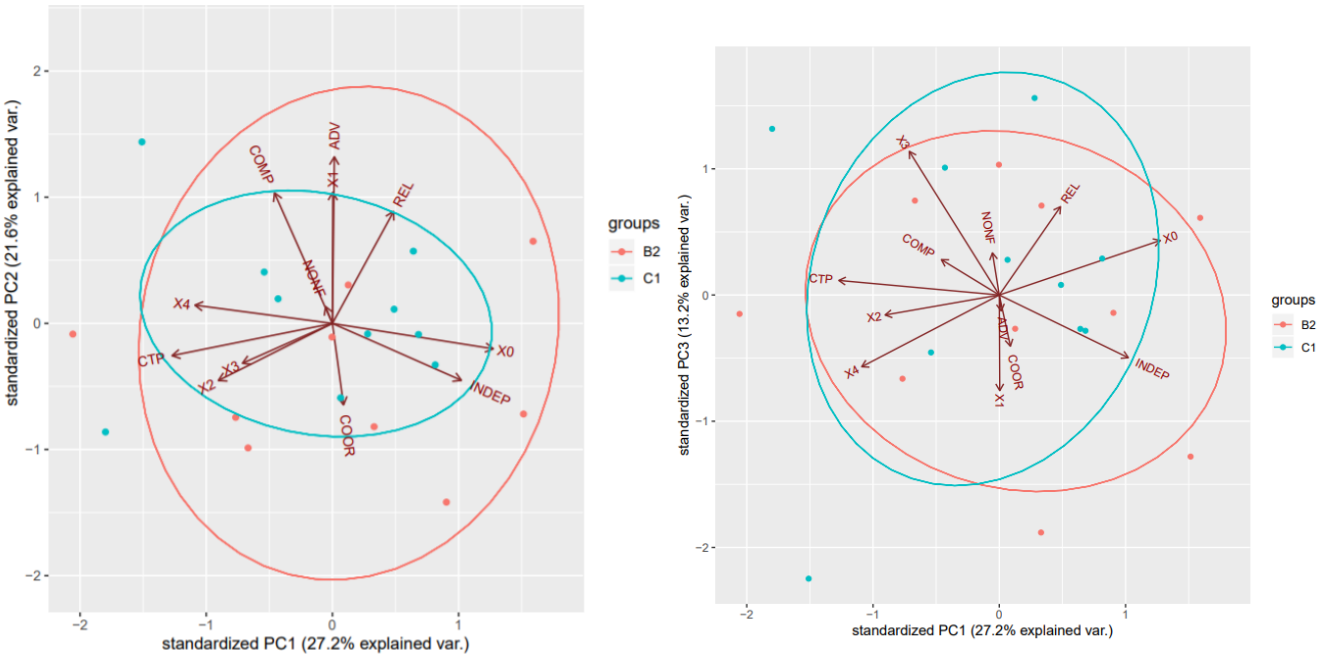
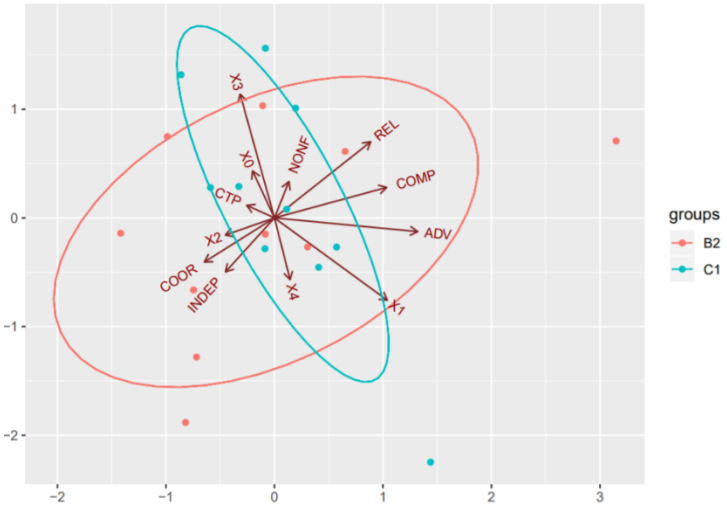



***Fig. 13–15*** *PCA biplots showing projections of data on subspace spanned by first three (1v2; 1v3; 2v3) PCA eigenvector*




Figures 13–15 show that structural complexity variables do not form any distinct cluster, which indicates that the dataset cannot be explained by a reduced number of variables.

## 6.3 Discussion

The quantitative analysis of productive and structural complexity showed that the datasets were very similar in the score distribution both with respect to both crude and fine-grained measures. Higher phrasal complexity is ascribed to advanced speakers in the research (Norris & Ortega 2009). MLC was, however, among the measures with the highest uniformity of score distribution. In fact, the between-group comparison of Vercellotti's (2018) complexity-scale levels revealed coincidental structural patterns shared among the participants: the largest proportion of all the monologues was comprised of simple independent clauses, which were then complemented with Level 3 & 4 structures and less frequently with Level 1 & 2 structures. This might come as surprising, as adverbial clauses of reason introduced by *because* are rather typical of spoken language (Foster et al. 2000). The number of such structures was, however, excluded from the analysis in the process of AS-unit coding to reduce their skewing effect. Speakers showed a tendency to mark the beginning of a new utterance with *because*, which, similarly to *and*, was used more as a filler rather than as a true conjunction.

The results of the analysis confirmed many of the concerns raised in this thesis pertaining to the limitations of the complexity construct and properties of spoken language (see Chapter 2.1.1 and Chapter 2.2.1.1)

1) It has been frequently argued that spoken syntactic complexity may be inhibited by limited attentional and cognitive resources (Skehan 1998). Vercellotti (2018) discovered in her developmental study that high-intermediate and low-advanced speakers produced more complex structures after intensive exposure to L2, and suggested that syntactic complexity might not be as largely inhibited by cognitive capacity as it had been claimed. This thesis, comparing upper-intermediate and advanced students, found no significant increase in complexification in the C1 group, and therefore limitations in cognitive load cannot be ruled out as one of factors potentially impeding syntactic complexity at higher proficiency levels.
2) Other, possibly most significant inhibiting factors lie at the very core of the complexity construct. Idiosyncrasy, deliberateness, and motivation have been reoccurring in the research as the reasons why higher complexity cannot equal better language production (Skehan 1998, Pallotti 2015, Michel 2017). Complexity is arguably the most flexible component of the CAF triad, and possibly the only one that can be shaped by the linguistic choices a speaker, who may feel motivated to produce language of varying complexity depending on the circumstances (Skehan 1998). Pallotti (2015) described intra-individual variation as stylistic dimension of syntactic complexity, which reflects "culture-specific rhetorical patterns" and may vary based on stylistic choices of each individual. Speakers may also feel limited by the conventions connected with the mode of production and the register. The monologic task, the syntactic properties of which were measured in this thesis, was accompanied by an instruction to follow an informal tone, which automatically

imposed sociocultural limits on the conversation. Having the listener in mind, they may feel obliged to leave out overtly complex structures that would overload the listener's cognitive resources.

3) With regards to spontaneous speech, the research has not agreed on its basic constituting elements. Biber (2011) identified subordinate clauses as typical for daily conversation. In contrast, Trebbits (2014) and Schiffrin (2014) claimed that the speech is too ephemeral to be produced in complex chunks of texts, and suggested short simple clauses as the building blocks of spoken language. The results of this analysis confirmed that out all clause types measured, independent clauses accounted for the largest portion of the total amount of clauses produced. At the same time, speakers made an extensive use of subordination, and preferred more complex structures to lower-level multi-clausal units, as the structural complexity counts revealed. Combined with the high dispersion of scores in the majority of measures, the results highlighted the variability within the sample rather than dominance of a single constituting structure.

### 6.3.1 Limitations of the Study: The Sample

LINDSEI_CZ, from which the sample derives, contains recordings of 36 C1 speakers and only 12 B2 speakers. To achieve mutual comparability of the subsamples, 10 speakers were selected from each group, which is a much lower sample size than in published L2 complexity studies (e.g. Boulté & Housen (2015) gathered their data from 94 participants; De Clercq & Housen (2017) from 581 participants (cross-sectional study); Vercellotti (2018) from 66 participants). At the same time, the relatively small sample size enabled us to perform a detailed analysis of structural complexity in individual speakers and provide a more detailed picture of within-group variation across the measures.

The second, and perhaps the most substantial limitation regarding the sample, is the proximity of the participants in terms of their language proficiency. Although ranked as B2 and C1 speakers, the participant achieved very similar scores in the holistic ratings. The inconclusiveness of the results presented in this study may then likely stem from the weak contrast in their proficiency scores. A much wider gap in language proficiency would likely reveal more noticeable differences between the groups. The analysis would therefore benefit from an additional sample consisting of higher- or lower-level participants from the same institutional background. This was, however, not possible, as only none of the speakers in LINDSEI_CZ was ranked below the B2 level and only two speakers were identified as C2 in the holistic ratings[23].

### 6.3.2 Limitations of the Study: Measure Validity

The measures employed in this study were chosen to reflect the complexity construct identified in section 3.2. As recommended by Lahmann et al. (2015), cruder length-based and subordination measures were combined with fine-grained indices of structural complexity to gain a deeper understating of the sources of complexification. The analysis of the data distribution confirmed that length-based and subordination measures were – despite their robustness and recurrence in L2 syntactic complexity research – highly opaque. For instance, the large-grained subordination measure (C/AS) was only informative when complemented by structural complexity metrics. Mean length of clause (MLC) was disputed by Boulté & Housen (2012) for its vagueness, but was retained in the analysis to enhance the replicability of the research. No further phrasal complexity metrics were added for several reasons. Firstly, in spoken complexity research, finer-grained indices suitable for very advanced learners have only been called for, yet not established. More specific phrasal complexity metrics may be available

[23] More on the ratings in Huang, Kubelec, Keng & Hsu (2018).

for the analysis of written data but are not common in the research of spoken syntactic complexity. At the same time, the participants were too advanced for the crude measures to be informative. For example, the frequency of coordinated verb phrases may be contrasted with subordination measures once the participants are in earlier stages of language development, where the transition from coordination to subordination is be more noticeable. In this study, participants produced only a very limited quantity of coordinated phrases and instead made use of simple, independent clauses combined with constellations of various types of subordinate clauses. One of the outcomes of the thesis is thus the confirmation of the need to establish more fine-grained indices suitable for very advanced speakers, which has already been expressed by Lahmann et al. (2015).

# 7 Conclusion

The aim of my thesis was to explore patterns of syntactic complexity in two groups of fairly advanced students and, perhaps more importantly, to bring the reader's attention to the fair amount of caveats that a researcher encounters when pursuing spoken language analysis. Operationalizing spoken syntactic complexity proved to be an unpredictable enterprise, where one must constantly make compromises to mediate between replicability and validity of the research. Foster et al. (2000) end their paper admitting that crucial decisions must be made when analysing spoken texts, which are left at the researchers' discretion. The final chapter of the thesis will identify such decisions together with limitations of L2 complexity as a theoretical construct and the missing links that are yet to be established in the research field.

## 7.1 Construct Validity

One of the controversies associated with complexity has been its polysemic nature, and more importantly, the failed attempts of many researchers to define an underlying theoretical complexity construct in their studies (Housen et al. 2012). Not relying on any constitutional theoretical framework, the CAF triad has dominated the SLA field with much freedom left in the hands of the researchers. Thus, mainly in the complexity research, many studies came into existence with only weak links between theory, operationalization, and interpretation. After influential publications of Ellis & Barkhuizen (2005), Norris & Ortega (2009), Pallotti (2009, 2015) and Housen et al. (2012), the focus has shifted towards producing a body of research that would be rid of all its former opacity, and where appropriate measures would lean on a robust theoretical basis. I believe that the underlying complexity construct was firmly established in this thesis based on the account of Housen et al. (2012), and that the measures employed in the research were selected accordingly. Syntactic complexity was here defined as the number and the nature of discrete syntactic structures that a speaker produces, and the number and the nature of relationships between them and their constituting elements. The battery of measures of spoken syntactic complexity employed in this study was chosen to meet the criterion of replicability, which was reached by using the most frequent length-based and subordination measures recommended by Norris & Ortega (2009), and at the same time answered the need for including more fine-grained indices of structural complexity.

## 7.2 Complexity and L2 Proficiency

The relationship between linguistic complexity and L2 proficiency has been a subject of many debates (Michel 2017). The link between the two constructs has not been discovered, and many critics are inclined to dispute that higher complexity could equal better language performance (Pallotti 2009). Syntactic complexity may rise linearly with task complexity, i.e.

when the speaker is instructed and motivated to produce complex language (Skehan 1998). However, in spontaneous speech, there appear too many confounding variables, stemming from the very nature of spoken language (more in Chapter 2.2), which contest the assumption that learner language complexifies with increasing proficiency. This thesis does not work with the hypothesis that syntactic complexity predicts language proficiency. The aims of the analysis were purely descriptive in this respect. It was carried out to discover whether syntactic complexity changes at different but still fairly advanced levels of proficiency and identify the trajectory and the nature of the change. The quantitative analysis showed no significant effect of proficiency level on productive and structural complexity. However, due to the small sample size and close proximity of language proficiency between the groups, no conclusive claims concerning the interaction between syntactic complexity and L2 proficiency may be inferred from the results. What the analysis did show was considerable within-group variation, especially in the lower-proficiency students, where individual speakers often obtained much higher complexity scores than any of their peers. The speaker who consistently attained outstanding complexity scores (see speaker B2_3 in Appendix 1, Appendix 2, and Appendix 3) was ranked as B2 in the holistic ratings.

## 7.3 Analysing Spoken Language

The secondary aim of the thesis was methodological – on the analysis of twenty monologues of advanced speakers of English with Czech as their L1, I attempted to identify weak or missing links in operationalization of spoken L2 complexity and the areas where there is still much room for improvement. Spoken language is notoriously difficult to analyse, and some issues may never be resolved due to conflicting theories on the most suitable unit of analysis. The dispute on whether clause or sentence is the basic unit of spoken language have been settled in favour of the clause, and dividing spoken texts into sentences is now generally frowned upon (Miller & Weinert 2009). In contrast, studies still vary in their definition of a

clause, which may (or may not) include non-finite verb phrases (compare e.g. Foster et al. (2000) and Lu (2010)).

On the multi-clausal level, Foster et al. (2000) made a substantial contribution towards the unification of spoken language analysis, designing the AS-unit, which has since its emergence become the staple of spoken complexity research within the CAF framework. On the one hand, the AS-unit coding instruction allows for the specifics of speech, such as its elliptic nature, pauses, fillers, or false starts (more in section 2.2.1.1.3). On the other hand, the process of coding is extremely laborious. The intonation and pause criteria, although necessary to analyse peculiar cases, make it impossible to fully automate the text division. Each coded transcription had to be thoroughly inspected several times, which made the whole process very time-consuming. Another disadvantage of the AS-unit is that the coding criteria do not include a clear definition of a word. To facilitate easier analysis of the texts, studies are unequivocal in understanding a word as a string of phonemes meaningful when uttered in isolation. As I outlined in section 2.2.1.1.3, this definition is problematic in that it privileges the speakers who use longer proper names or memorized sequences of text (such as idiomatic expressions). While the idea of establishing a universally applicable definition is debatable, inventing fine-grained indices that would neutralize such kind of variation is certainly feasible.

Data exclusion was the process where, owing to the nature of the task, there was not much room for compromises. Foster et al. (2000) designed a three-tiered system of application, which reflects the CAF component analysed as well as the nature of the data. A very large portion of the texts was excluded to achieve mutual comparability of the samples. No dialogic exchanges could be consequently analysed, as the responses of the speakers were elliptical to various extent. I am aware that many issues disappeared with the significant amount of data being excluded from the analysis, and that dialogic or otherwise very elliptical texts pose much

more challenges to the analysts. Nonetheless, the strict criteria were established to enhance the validity of the syntactic complexity measures and the underlying complexity construct.

Finally, in section 2.2.1.1.3 I proposed establishing a clear typology of AS-units based on their length and structure as one of possible solutions towards more transparent operationalization of L2 syntactic complexity. A question then arises as to whether such a typology would not be too restrictive for the researchers, who build their studies on different theories and complexity constructs. And if such typology should come into existence, how many structures should it encompass in order to be universally applicable? Perhaps the fact that researchers must make the crucial decisions themselves is the only viable alternative. It is then vital that studies start sharing descriptions of their methods.

## 7.4 Automated Analysis of Syntactic Complexity

A number of issues in operationalization of written L2 complexity were resolved with the emergence of automated tools. The source of the former scarcity of measures employed across the complexity studies were, according to Lu (2010), a result of laboriousness of manual computation. The tools available before 2010 only analysed complexity measures that were not typically used in current L2 complexity studies. Informed by the survey of syntactic complexity measures by Wolfe-Quintero et al. (1998), Lu (2010) selected fourteen of the most frequent and robust units and designed L2 syntactic complexity analyser, which enabled for processing of a high number of texts and analysing several measures at the same time. Kyle (2016) and his TAASC complemented L2SCA metrics with an array of fine-grained indices. In contrast, no automated tools are available to analyse spoken syntactic complexity.

Both L2SCA and TAASC use punctuation criteria to divide the text into the units of analysis (T-unit, sentence, clause). Dividing transcribed spoken texts into AS-units cannot be fully automated, since the process requires the use of the recording to apply tone and pause

criteria. Moreover, measures generally available for gauging written complexity may not be as valid in the spoken context, and more fine-grained indices, e.g. those of structural complexity, are rather rare in studies on spoken syntactic complexity. There is also a great variability in the measures employed in spoken complexity research, and no attempt to create a set of established syntactic complexity measures similar to Lu's (2010) has yet been made. The thesis thus attempted to compare manual calculations of MLC, MLU, and C/As with scores obtained from L2SCA. Due to the conflicting definitions of the clause, AS-unit coding, even though the monologic and therefore heavily pruned texts were adapted to correspond with the L2SCA text format, is not compatible with T-unit coding of the tool, which proved that syntactic complexity of spoken texts cannot be accurately gauged using automated analysers.

## 7.5 Variation in Syntactic Complexity

A final thought is dedicated to the phenomenon that has dominated the SLA research in the recent years (Kuiken et al. 2019). Idiosyncrasy, deliberateness, and motivation were introduced earlier as possible confounding variables that may have a skewing effect on the complexity scores. In reality, they seriously undermine the current conception of L2 complexity and the ways in which the construct is operationalized. Linguistic complexity depends, unlike accuracy and fluency, on the choices that the speaker makes based on the circumstances. All the three notions are integral parts of intra-speaker variation and depend upon non-linguistic factors.

Variation is, in fact, at the heart of any language performance. A single speaker will deliberately choose different syntactic structures on different occasions. At the same time, two speakers will very unlikely express the same thought using the same linguistic means. Kuiken et al. (2019) have identified four sources of variation in syntactic complexity that has been given attention in current research: developmental variation, variation in task modality (spoken versus

written language), variation in task type and genre, and variation in source and target language. At the same time, Pallotti (2015) speaks of stylistic syntactic complexity, which is, unlike grammatical complexity, always dependent on the speaker's deliberate choices. But how do we measure the degree of deliberateness and motivation in spontaneous spoken production? If we asked the speakers about their syntactic choices, would they be able to explain why they used certain structures more than others? Most likely not. Task-based research showed that a task instruction might motivate the speaker to produce highly complex language, but a part of the performance will always remain in the speaker' hands. More attention should be, therefore, paid to in-depth study of individual language development in order to discover more about the role of intra-individual variation on syntactic complexity of spoken production.

# 8 Bibliography

Albert, Á. and Kormos, J. (2011). Creativity and Narrative Task Performance: An Exploratory Study. *Language Learning*, 61, 73-99.

Bajaj, A. (2007). Analysis of oral narratives of children who stutter and their fluent peers: Kindergarten through second grade. *Clinical Linguistics & Phonetics*, *21*(3), 227–245.

Biber, D., Gray, B., & Poonpon, K. (2011). Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development? *TESOL Quarterly*, 45(1), 5–35.

Bialystok, E. (1994). Analysis and Control in the Development of Second Language Proficiency. Studies in *Second Language Acquisition*, 16(2), 157–168.

Boulté, B. and Housen, A. (2012). Defining and operationalising L2 complexity. In: A. Housen, F. Kuiken and I. Vedder, ed., *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA.* Amsterdam: John Benjamins.

Brumfit, C. J. (1984). *Communicative Methodology in Language Teaching: The Roles of Fluency and Accuracy.* Cambridge: Cambridge University Press.

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics*, *30*(7), 489–518.

Bryfonski, L., & Sanz, C. (2018). Opportunities for Corrective Feedback During Study Abroad: A Mixed Methods Approach. *Annual Review of Applied Linguistics*, *38*, 1–32.

Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1(1), 1–47.

Connor, U., & Schneider, M. (1990). Analyzing topical structure in ESL essays. *Studies* in *Second Language Acquisition*, *12*(04), 411–427. Dumanig, F. P.

Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (p. 83–113). Academic Press.

Chaudron, C. (1988). *Second Language Classrooms: Research on Teaching and Learning.* Cambridge: Cambridge University Press.

Council of Europe. (2001). *Common European Framework of Reference for Languages. Learning, Teaching, Assessment.* Cambridge University Press.

Cvrček, V., & Richterová, O. (eds). "cnk:lindsei_cz". *Příručka ČNK*

Cambridge University Press (2020). *Cambridge online dictionary.* Cambridge dictionary online.

De Clercq, B., & Housen, A. (2017). A Cross-Linguistic Perspective on Syntactic Complexity in L2 Development: Syntactic Elaboration and Diversity. *The Modern Language Journal*, 101(2), 315–334

Ellis, N. C. (1996). Sequencing in SLA. *Studies in Second Language Acquisition*, 18(1), 91–126. Cambridge University Press.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford University Press.

Fukuta, J. (2015). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20(3), 321-340.

Foster, P., Tonkyn, A. and Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. Applied Linguistics, 21(3), 354-375.

Freeborn, D., French, P., & Langford, D. (1993). Spoken English and written English. In D. Freeborn, P. French, & D. Langford, *Varieties of English* (pp. 76–103). Macmillan Education UK.

Gráf, T. (2015). Accuracy and Fluency in the Speech of the Advanced Learner in English.

Univerzita Karlova, Filozofická fakulta.

Gráf, Tomáš. (2017). LINDSEI_CZ: A Corpus of Spontaneous Spoken English of Advanced Speakers. *Institute of the Czech National Corpus FF UK*. (https://wiki.korpus.cz/doku.php:en:cnk:linsei_cz).

Gaëtanelle, G., De Cock, S., & Granger, S (2010). *The Louvain International Database of Spoken English Interlanguage*. Handbook and CD-ROM. Louvain-la-Neuve: Presses universitaires de Louvain.

Hakuta, K. (1975). Becoming bilingual at age five: the story of Uguisu. Unpublished thesis, Harvard University.

Halliday, M., & Matthiessen, C. (2004). *An introduction to functional grammar* (2nd ed.). London: Arnold.

Housen, A., & Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), 461–473. Oxford University Press (OUP).

Huang, L., Kubelec, S., Keng, N., & Hsu, L. (2018). Evaluating CEFR rater performance through the analysis of spoken learner corpora. *Language Testing in Asia* 8(1). 14.

Hunt, K. (1965). Grammatical Structures Written at Three Grade Levels. Champaign,Ill: National Council of Teachers of English.

Hunt, K. (1966). Recent measures in syntactic development. *Elementary English*, 43, 732–139.

Hunt, K. (1970). Syntactic Maturity in School-children and Adults. Monograph of the Society for Research into Child Development.

Karayayla, T., & Schmid, M. S. (2018). First Language Attrition as a Function of Age at Onset of Bilingualism: First Language Attainment of Turkish-English Bilinguals in the United Kingdom. *Language Learning*.

Kuiken, F. & Vedder, Ineke & Gilabert, Roger. (2010). Communicative adequacy and

linguistic complexity in L2 writing. Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research. Eurosla monographs.

Kuiken, F. and Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. *Dimensions of L2 Performance and Proficiency*, 143 – 170.

Kuiken, F., Vedder, I., Housen, A., & De Clercq, B. (2019). Variation in syntactic complexity: Introduction. *International Journal of Applied Linguistics*.

Kramsch, C. (1986). From Language Proficiency to Interactional Competence. *The Modern Language Journal*, 70(4), 366–372.

Kyle, Kristopher. (2016). Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-Based Indices of Syntactic Sophistication. Dissertation. Georgia State University.

Kyle, K. and Crossley, S. (2018). Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices. The Modern Language Journal, 102(2), 333-349.

Lambert, C., & Nakamura, S. (2018). Proficiency-related variation in syntactic complexity: A study of English L1 and L2 oral descriptive discourse. International Journal of Applied Linguistics, 29(2), 248–264.

Lahmann, C., Steinkrauss, R., & Schmid, M. S. (2015). Factors Affecting Grammatical and Lexical Complexity of Long-Term L2 Speakers' Oral Proficiency. Language Learning, 66(2).

Lahmann, C., Steinkrauss, R. and Schmid, M. (2019). Measuring linguistic complexity in long-term L2 speakers of English and L1 Attriters of German. *International Journal of Applied Linguistics*

Lambert, C. and Engler, S. (2006). Information Distribution and Goal Orientation in Second Language Task Design. *Investigating Tasks in Formal Language Learning*, 27-43.

Larsen-Freeman, D. (1978). An ESL Index of Development. *TESOL Quarterly*, 12(4), 439.

Larsen-Freeman, D. (2009). Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics,* 30(4), 579–589. Oxford University Press.

Larsen-Freeman, D., & Strom, V. (1977). The Construction of a Second Language Acquisition Index of Development. *Language Learning*, *27*(1), 123–134.

Lázaro-Ibarrola, A., & Hidalgo, M. Á. (2017). Procedural repetition in task-based interaction among young EFL learners. *ITL - International Journal of Applied Linguistics*, 168(2), 183–202.

Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach. *Language Learning*, 40(3), 387-417.

Levkina, M. and Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. *Dimensions of L2 Performance and Proficiency*, 171-198. Champaign, Ill: National Council of Teachers of English.

Loban, W. (1966). *The Language of Elementary School Children* (Research Report no. 1)

Loban W. (1976). Language development: Kindergarten through grade twelve. Urbana, IL: National Council of Teachers of English.

Lu, X. (2010). A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. *TESOL Quarterly*, *45*(1), 36–62.

MacWhinney, B. (2020). Tools for Analyzing Talk. Part 1: The CHAT Transcription Format. Carnegie Mellon University.

Magne, V., Suzuki, S., Suzukida, Y., Ilkan, M., Tran, M., & Saito, K. (2019). Exploring the Dynamic Nature of Second Language Listeners' Perceived Fluency: A Mixed-Methods Approach. *TESOL Quarterly.*

Mäkinen, L., Loukusa, S., Laukkanen, P., Leinonen, E., & Kunnari, S. (2014). Linguistic and pragmatic aspects of narration in Finnish typically developing children and children with specific language impairment. *Clinical Linguistics & Phonetics*, *28*(6), 413–427.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing Linguistic Features of Cohesion. *Discourse Processes,* 47(4), 292–330.

Michel, M. (2017). Complexity, Accuracy, and Fluency in L2 Production. In L. Shawn & S. Masatoshi (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition*. Routledge.

Michel, M., Kuiken, F. and Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL - International Review of Applied Linguistics in Language Teaching*, 45(3).

Miestamo, M. (2008). Grammatical complexity in a cross-linguistic perspective. In *Language complexity: Typology, contact, change* (Vol. 2008). John Benjamins Pub. Co.

Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.). (2008). *Language complexity: Typology, contact, change*. John Benjamins Pub. Co.

Miller, J.E., and R. Weinert. (2009). *Spontaneous Spoken Language: Syntax and Discourse*. Oxford; New York: Oxford University Press.

Moser, J. (2010). Using an AS-unit complexity benchmark to measure beginner learner oral production in communicative tasks. Osaka Shoin Women's University, Faculty of Liberal Arts, collected essays (47), 191-201.

Nippold, M. A., Vigeland, L. M., & Frantz-Kaspar, M. W. (2017). Metacognitive verb

production in adolescents: The link to complex syntax. *Clinical Linguistics & Phonetics*, *31*(10).

Norris, J. and Ortega, L. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics*, 30(4), 555-578.

Pallotti, G. (2009). CAF: Defining, Refining and Differentiating Constructs. *Applied Linguistics*, 30(4), 590–601.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117-134.

Pica, T. et al. (1989). Comprehensible outputs as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*. 11(1): 63–69.

Quirk, R. (1985). A Comprehensive Grammar of the English Language. London: Longman.

Rescher, N. (1998). *Complexity. A Philosophical Overview.* Transaction Publishers.

Révész, A., Ekiert, M., & Torgersen, E. N. (2014). The Effects of Complexity, Accuracy, and Fluency on Communicative Adequacy in Oral Task Performance. *Applied Linguistics*.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA. In *Cognition and Second Language Instruction* (pp. 287–318). Cambridge University Press.

Santos, S. (2018). Effects of task complexity on the oral production of Chinese learners of Portuguese as a foreign language. *Journal of the European Second Language Association,* 2(1), 49–62.

Schiffrin, D. (2014). Discourse. In Fasold, RW & Connor-Linton, J. (Eds.). *An Introduction to Languages and Linguistics*. Cambridge: Cambridge University Press.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1–14. Cambridge

University Press (CUP).

Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, *30*(4), 510–532.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. Language Teaching Research, 1(3), 185–211.

Skehan, P., & Foster, P. (2008). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, & I. Vedder (Eds.). *Complexity, accuracy and fluency in second language use, learning and teaching*, 201–222. Wetteren: Universa Press.

Takamasa, I, Maeda, R., Ogawa, K., Yoshikawa, Y., Ishiguro, H., Suzuki, K., Aoki, T., Maesaki, M. and Hama, M. (2018). Improvement of Japanese adults' English speaking skills via experiences speaking to a robot. *Journal of Computer Assisted Learning*, 35(2), 228-245.

Tavakoli, P. and Foster, P. (2008). Task Design and Second Language Performance: The Effect of Narrative Type on Learner Output. *Language Learning*, 58(2), 439-473.

Tavakoli, P. and Skehan, P. (2005). 9. Strategic planning, task structure and performance testing. *Language Learning & Language Teaching*, 239-273.

Vachek, J., and Luelsdorff, P.A. (1989). *Written Language Revisited*. Amsterdam/Philadelphia: John Benjamins Pub. Co..

Vercellotti, M. L., & Packer, J. (2016). Shifting structural complexity: The production of clause types in speeches given by English for academic purposes students. *Journal of English for Academic Purposes*, 22, 179–190.

Vercellotti, M. L. (2019). Finding variation: Assessing the development of syntactic complexity in ESL Speech. *International Journal of Applied Linguistics*, *29*(2), 233–

247.

Wolfe-Quintero, K., Inagaki, S., and Kim, H.-Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, & Complexity.* Honolulu: Second Language Teaching & Curriculum Center, University of Hawaii.

Young, R. and Milanovic, M. (1992). Discourse Variation in Oral Proficiency Interviews. *Studies in Second Language Acquisition*, 14(04).

# 9 Resumé

Otázka, jak definovat pokročilého mluvčího, figuruje již po několik desetiletí ve výzkumu akvizice druhého jazyka a dala vzniknout mnoha modelům jazykové kompetence (např. dnes tradičně využívanému modelu čtyř dovedností, který tvoří čtení, psaní, mluvení a poslech). S cílem kvantitativně změřit jazykový projev a pokročilost nerodilých mluvčích vznikl jazykový model sestávající ze tří komponent: komplexnosti, přesnosti a plynulosti jazykového projevu (anglicky Complexity, Accuracy, Fluency, zkráceně CAF) (Housen 2012). Tato práce zabývá syntaktickou komplexností mluveného projevu nerodilých mluvčích angličtiny na úrovni B2 a C1, jejichž mateřským jazykem je čeština. Komplexnost je zde definována jako „množství a povaha komponent, ze kterých se určitá jazyková struktura skládá, a množství a povaha vztahů mezi jejími konstituenty (přeloženo z Housen et al. 2012., str. 22)

Drtivá většina kritik se shoduje na tom, že právě komplexnost je nejproblematičtější dimenzí triády CAF, a to především z toho důvodu, že je jako teoretický koncept polysémní (Pallotti 2015). Zahrnuje tedy řadu odlišných dimenzí (např. subjektivní *složitost* určitých jazykových jevů z pohledu mluvčího), které pak studie často nedokáží řádně definovat. Model CAF se totiž nezakládá na konkrétní teorii o vývoji druhého jazyka, která by tři dimenze sjednotila – přesnost a plynulost se zkoumaly v kontextu jazykové třídy již v 80. letech 20. století, ale součástí modelu jazykové kompetence se staly až o dekádu později, kdy je Skehan doplnil o komplexnost (Housen et al. 2012).

Za dobu své existence se staly tyto tři dimenze základními měřitelnými komponenty žákovského jazyka. Komplexnost jazykové produkce se analyzuje na rovině syntaxe, lexikologie, fonologie a morfologie, přičemž syntaktická dimenze je jednoznačně ze všech nejprobádanější. Výzkum se však věnoval především měření psaného jazyka, zatímco mluvený projev byl studiemi opomíjen. Operacionalizace mluvené syntaktické komplexnosti je totiž

značně pracnější. Řeč je charakteristická vysokou mírou eliptických promluv, vyplněných pauz, falešných startů a opravení vlastní chyby mluvčími. Přepisy mluvených nahrávek pak je potřeba segmentovat do syntaktických jednotek, na kterých se následně komplexnost měří.

V psaném jazyce se běžně analyzuje klauze a souvětí, které lze velmi spolehlivě vymezit za pomoci interpunkce. Nejčastější analytickou syntaktickou jednotkou nad úrovní klauze se však stala tzv. T-unit (tzn. nejkratší gramaticky správná věta (sentence), na kterou lze rozdělit psaný jazyk) (Hunt 1965). Ta ve své původní podobě sestávala z věty hlavní a všech vět vedlejších. Později došlo k adaptaci definice pro analýzu mluveného jazyka, kde T-unit tvořila věta hlavní a klauze k ní přidružené. Jednotka však i přesto nereflektovala specifika mluveného jazyka, což vedlo ke značným nepřesnostem v měření. Loban (1966) se pokusil T-unit přetvořit tak, aby vedla k přesnější analýze. Samostatnou jednotkou se mohly tak stát i fráze, jednotlivá slova či gramaticky nesprávné věty v momentě, kdy je bylo možné na základě předchozího kontextu rozvést do úplných klauzí. V roce 2000 pak Foster et al. identifikovali znaky řeči, které všechny dosavadní jednotky opomíjely, a vyvinuli na jejich základě vlastní AS-unit (jednotku analýzy řeči), kterou doplnili o rozsáhlá kódovací kritéria s instrukcemi k segmentaci transkriptů.

I přes detailní popis kódování AS-units identifikovala tato práce šedé zóny, jejichž vyřešení nechává studie Foster et al. (2000) na uvážení výzkumníků. Nenavrhuje například vlastní definici slova. V drtivé většině případů se ve studiích jedním slovem rozumí řada fonémů, které dávají smysl i v izolaci. Tato definice je nejpraktičtější z hlediska automatické analýzy textů, zvýhodňuje však mluvčí, kteří používají ve svých promluvách dlouhá vlastní jména nebo memorované (většinou idiomatické) sekvence slov. Je značně nepravděpodobné, že by přišla v platnost definice, kterou by bylo možné aplikovat ve všech kontextech CAF. Rozdíly v datech, které dosavadní definice způsobují, je však možné zmírnit detailnějším měřením syntaktických struktur a vztahů mezi nimi.

Další problematickou oblastí v analýze mluvených textů je rozsah dat, která se z analýzy vynechávají podle povahy výzkumu a dimenze modelu CAF, která je předmětem analýzy. Vyplněné pauzy jsou například zásadní pro měření plynulosti projevu, nicméně při zkoumání jakékoli dimenze lingvistické komplexnosti mohou značně zkreslit výsledky analýzy. Foster et al. (2000) navrhují třístupňový systém vyřazování dat v závislosti na povaze vlastní analýzy: první stupeň se používá v momentě, kdy je potřeba text zachovat v plném rozsahu, a neodstraňují se tedy žádná data. Druhý stupeň je vhodný např. pro dialogické, značně eliptické texty, kde Foster et al. (ibid.) doporučují vyřadit pouze jednoslovné promluvy či doslovné odpovědi (přesné opakování slyšeného). Poslední, třetí stupeň vyřazování dat, Foster et al. (ibid.) doporučují použít pouze ve výjimečných případech, kdy je žádoucí zachovat pokud možno co nejcelistvější syntaktické jednotky. Z analýzy se tedy vynechávají kromě dat z druhé úrovně také eliptické promluvy a neslovesné věty. Tříúrovňový systém však nezohledňuje poznámky mluvčího adresované tazateli nebo komentující průběh rozhovoru. Lennon (1990) je nazývá „asides“, tzn. výpovědi vyřčené mimo zadanou úlohu. V momentě, kdy studie pracují s AS-units, je záhodno, aby zveřejnily rozsah odstraněných dat pro jejich případnou replikaci. Součástí této práce byl tedy průzkum, kde bylo na dvaceti studiích publikovaných od roku 2000 ukázáno, že jejich drtivá většina pouze odkazuje na definici jednotky podle Foster et al. (2000) a jejich systém odstranění dat. Foster et al. (2000) přiznávají, že na taková data často v publikacích nezbývá místo. Pokud ale výzkumníci nezačnou zveřejňovat podpůrné informace k práci s daty, ovlivní tak negativně replikovatelnost svých studií.

Bez chyb nejsou ani způsoby, jakými se samotná syntaktická komplexnost měří. Boulté & Housen (2012) ve svém průzkumu čtyřiceti studií komplexnosti žákovského jazyka zjistili, že z několika desítek dostupných metrik syntaktické komplexnosti figuruje pouze velmi malá část v téměř všech studiích, zatímco drtivá většina z nich se vyskytuje pouze v počtech jednotek. Norris & Ortega (2009) upozornili na to, že validita komplexnosti jako teoretického

konstruktu bývá ve studiích velmi nízká právě z toho důvodu, že zvolené metriky neměří jimi nadefinovanou syntaktickou komplexnost v jejím plném rozsahu. Často se stává, že studie využívají redundantních metrik místo toho, aby je nahradily detailnějšími indikátory syntaktické komplexnosti, které jsou v analýze více potřeba. Ve výzkumu značně převažují metriky *produktivní komplexnosti***,** založené na průměrné délce vícevětné jednotky (např. T-unit a AS-unit), které měří tzv. „obecnou komplexnost", a na podílu vedlejších vět (metriky subordinace) (ibid.), měřící tzv. „klauzální komplexnost" (ibid.). Pro úplnost měření navrhují Norris & Ortega (ibid.) přidat dimenzi frázovou, která se měří v podobě průměrné délky klauze. Metriky založené na délce syntaktické jednotky a podílu subordinace jsou ale značně vágní, jelikož neukazují zdroj komplexifikace. Mnohé kritiky (např. Lahmann et al. 2015, Kyle 2016) proto doporučují takové metriky komplementovat *strukturními* indikátory syntaktické komplexnosti (většinou měřené podílem specifických struktur v textu), díky nimž získáme celistvý obraz syntaktické komplexnosti projevu mluvčího.

V kontextu mluveného jazyka však nebyly strukturní indikátory dosud příliš prozkoumány. Zároveň je operacionalizace mluvené komplexnosti problematická i z toho důvodu, že zatím neexistují nástroje, které by proces měření zautomatizovaly. Pro psaný jazyk vznikl v roce 2010 nástroj L2SCA (L2 Syntactic Complexity Analyser; Lu 2010), který měří čtrnáct nejrobustnějších metrik syntaktické komplexnosti psaných textů. Zároveň texty automaticky dělí na syntaktické jednotky (T-unit, věta, klauze), což není v mluveném kontextu vzhledem k nutnosti použití zvukových nahrávek proveditelné. V roce 2016 následoval nástroj TAASC (Tool for the Automatic Analysis of Ayntactic complexity; Kyle 2016), který kromě metrik použitých v L2SCA nabízí i několik desítek specifických indikátorů komplexnosti klauzí a frází. Jedním z původních cílů této práce bylo zjistit, zda tyto nástroje lze použít pro měření komplexnosti mluvených textů v případě, že se jedná o velmi celistvé, monologické texty, a že jsou přepisy upraveny tak, aby fyzicky odpovídaly psaným textům. Ukázalo se však, že Foster

et al. (2000), podle kterých byly texty v této práci děleny do AS-units, se liší od Lu (2010) a Kyle (2016) v definici klauze. Definice Foster et al. (2000) zahrnuje na rozdíl od Lu (2010) a Kyle (2016) nefinitní slovesné fráze, což znamená, že výsledky jakékoli měření v jednom z nástrojů, které zahrnuje počet klauzí nebo jejich délku, nebudou odpovídat výsledkům manuálního měření. Pro potvrzení nekompatibility byly vybrány tři metriky (průměrná délka klauze, průměrná délka AS-unit a počet klauzí na AS-unit), které byly následně měřeny manuálně a pomocí L2SCA. Nízká míra korelace manuálních a automatických výpočtů (průměrná délka jednotky r = .76; průměrná délka klauze r = .39; klauze na AS-unit r = .69) nekompatibilitu potvrdila, a komplexnost tedy dále byla analyzována pouze poloautomaticky pomocí počítačových skriptů.

Diplomová práce si tedy kladla za cíl zodpovědět následující výzkumné otázky:

1) Existuje rozdíl v *produktivní syntaktické komplexnosti* nerodilých mluvčích anglického jazyka na úrovni B2 a C1, jejichž mateřským jazykem je čeština?
2) Existuje rozdíl v *strukturní syntaktické komplexnosti* nerodilých mluvčích anglického jazyka na úrovni B2 a C1, jejichž mateřským jazykem je čeština?

Mluvené texty pro tento výzkum pocházejí z korpusu LINDSEI_CZ (Gráf 2017), který vznikl v rámci mezinárodního mluveného korpusu LINDSEI pod záštitou Centra pro anglickou korpusovou lingvistiku při Katolické univerzitě v Lovani (Cvrček a Richterová 2017). K analýze syntaktické komplexnosti byl vybrána jedna ze tří mluvených úloh, které tvořily každou z nahrávek. Mluvčí si měli vybrat jedno ze tří témat a samostatně o něm pak hovořit po dobu 3–5 minut. Jednalo se tedy o monology. Texty byly rozděleny do AS-unit na základě instrukce Foster et al. (2000). Pro získání co nejcelistvějších jednotek byla odstraněna nežádoucí data podle stupně 3 z Foster et al. (2000). Práce rovněž uvádí ošetření problematických případů jak v dělení jednotek, tak odstraňování dat. Produktivní metriky syntaktické komplexnosti byly zvoleny tak, aby odpovídaly třem úrovním komplexnosti

uvedených v Norris & Ortega (2009). Jednalo se o průměrnou délku AS-unit, průměrnou délku klauze, a průměrný počet vět na jednu AS-unit. K měření *strukturní komplexnosti* byla zvolena vážená pětistupňová škála strukturní komplexnosti (Vercellotti 2018), která popisuje struktury od jednoduchých samostatných vět (klauzí; úroveň 0) po komplexní vícevětné struktury sestávající z hlavní věty a minimálně dvou druhů vedlejších vět (úroveň 4). Zároveň byly texty kódovány podle větné typologie na základě Vercellotti & Packer (2016). Identifikovány byly vedlejší věty příslovečné, doplňkové (complement clauses), vztažné, poté souřadně spojené slovesné fráze a nefinitní věty. Zároveň byl změřen poměr struktur uvozených *I think, I guess*, protože nejsou oproti jiným strukturám obsahujícím doplňkové věty tolik komplexní. Strukturní komplexnost byla pak zjištěna pomocí výpočtu poměru struktur z každého stupně vážené škály v celkovém počtu struktur a poměrem typů vět k celkovému počtu vět.

Výsledky kvantitativní analýzy ukázaly, že mezi mluvčími na úrovni B2 a C1 podle SERR (Council of Europe 2001) není signifikantní rozdíl v žádné z měřených dimenzí produktivní a strukturní syntaktické komplexnosti. Ve skutečnosti byla distribuce dat velmi podobná u každé z měřených metrik. Odhalilo se, že bez ohledu na jazykovou úroveň tvořily největší část monologů jednoduché věty. Ty pak mluvčí ale nejčastěji kombinovali se strukturami 3. a 4. stupně vážené škály podle Vercellotti (2018), z čehož vyplývá, že všichni mluvčí produkovali vysoce komplexní jazyk nezávisle na jazykové úrovni. V rozložení dat byl ale často patrný vysoký rozptyl. Velké množství odlehlých hodnot zároveň poukázalo na rozdíly mezi jednotlivými mluvčími spíš než na skupinový trend.

K výzkumu se váže několik omezení, která mohla mít vliv na výsledek analýzy. Vzorek dvaceti mluvených textů je oproti počtu mluvčích v publikovaných studiích poměrně malý. Mluvčí byli rovněž na velmi podobné úrovni pokročilosti. Pro takto pokročilé mluvčí zároveň nebyly navrženy specifické indikátory syntaktické komplexnosti, které by fungovaly v mluveném kontextu (Lahmann et al. 2015). Je tedy možné, že i z tohoto důvodu nebyly mezi

skupinami patrné strukturní rozdíly. Ačkoli nebyl mezi skupinami zjištěn signifikantní rozdíl ani v produktivní, ani v strukturní syntaktické komplexnosti projevů mluvčích na úrovni B2 a C1, práce je přínosná metodologicky, jelikož poukázala na oblasti ve výzkumu mluvené syntaktické komplexnosti a analýzy mluveného jazyka, ve kterých je stále prostor pro zvýšení validity použitých metrik a replikovatelnost studií. Zároveň byly v práci identifikovány problematické aspekty samotného teoretického konstruktu komplexnosti v oblasti měření žákovského jazyka.

# 10 Appendix 1 – Productive Complexity Scores for Each Participant

| Participant | C1_1 | C1_2 | C1_3 | C1_4 | C1_5 | C1_6 | C1_7 | C1_8 | C1_9 | C1_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| C/AS | 1 .47 | 1 .75 | 1 .50 | 1 .48 | 2 .25 | 1 .50 | 1 .89 | 1 .95 | 1 .67 | 1 .65 |
| MLU | 9 .53 | 10 .88 | 10 .70 | 9 .48 | 16 .92 | 10 .70 | 11 .89 | 11 .61 | 12 .16 | 10 .94 |
| MLC | 6 .49 | 6 .22 | 7 .13 | 6 .41 | 7 .52 | 7 .13 | 6 .29 | 5 .94 | 7 .27 | 6 .64 |
| **Participant** | **B2_1** | **B2_2** | **B2_3** | **B2_4** | **B2_5** | **B2_6** | **B2_7** | **B2_8** | **B2_9** | **B2_10** |
| C/AS | 1 .25 | 1 .60 | 2 .32 | 1 .59 | 1 .40 | 1 .36 | 1 .54 | 1 .74 | 1 .96 | 1 .71 |
| MLU | 7 .82 | 11 .88 | 16 .07 | 11 .41 | 9 .87 | 8 .03 | 10 .76 | 11 .40 | 11 .54 | 11 .21 |
| MLC | 6 .27 | 7 .43 | 6 .92 | 7 .17 | 7 .06 | 5 .91 | 6 .97 | 6 .55 | 5 .89 | 6 .55 |

# 11 Appendix 2 – Proportion Scores for Individual Clause Types

| Participant | C1_1 | C1_2 | C1_3 | C1_4 | C1_5 | C1_6 | C1_7 | C1_8 | C1_9 | C1_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ADV | .05 | .05 | .12 | .10 | .17 | .06 | .07 | .11 | .11 | .04 |
| NONF | .03 | .14 | .03 | .04 | .11 | .06 | .12 | .12 | .06 | .05 |
| REL | .10 | .05 | .12 | .05 | .06 | .05 | .13 | .10 | .09 | .16 |
| COMP | .10 | .17 | .05 | .12 | .17 | .15 | .13 | .14 | .14 | .12 |
| COOR | .00 | .00 | .00 | .01 | .02 | .01 | .00 | .00 | .00 | .01 |
| INDEP | .50 | .26 | .40 | .47 | .15 | .43 | .24 | .26 | .40 | .37 |
| CTP | .03 | .14 | .00 | .01 | .09 | .05 | .03 | .02 | .00 | .03 |
| **Participant** | **B2_1** | **B2_2** | **B2_3** | **B2_4** | **B2_5** | **B2_6** | **B2_7** | **B2_8** | **B2_9** | **B2_10** |
| ADV | .05 | .08 | .20 | .10 | .08 | .04 | .06 | .07 | .13 | .06 |
| NONF | .02 | .20 | .04 | .06 | .05 | .08 | .11 | .14 | .09 | .08 |
| REL | .01 | .13 | .09 | .05 | .07 | .03 | .07 | .07 | .06 | .11 |
| COMP | .08 | .00 | .23 | .13 | .05 | .11 | .04 | .12 | .18 | .15 |
| COOR | .00 | .00 | .00 | .00 | .05 | .00 | .08 | .04 | .01 | .01 |
| INDEP | .77 | .35 | .16 | .40 | .53 | .53 | .43 | .30 | .26 | .31 |
| CTP | .00 | .00 | .00 | .03 | .00 | .06 | .00 | .09 | .09 | .00 |

# 12 Appendix 3 – Proportion Scores for Weighted Complexity Scale

| Participant | C1_1 | C1_2 | C1_3 | C1_4 | C1_5 | C1_6 | C1_7 | C1_8 | C1_9 | C1_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Level 0 | .72 | .47 | .60 | .68 | .35 | .65 | .52 | .50 | .65 | .61 |
| Level 1 | .05 | .00 | .13 | .08 | .21 | .05 | .06 | .07 | .04 | .06 |
| Level 2 | .03 | .12 | .02 | .02 | .06 | .09 | .06 | .10 | .04 | .03 |
| Level 3 | .15 | .27 | .17 | .16 | .09 | .18 | .23 | .12 | .14 | .27 |
| Level 4 | .05 | .14 | .08 | .07 | .29 | .02 | .13 | .21 | .12 | .03 |
| **Participant** | **B2_1** | **B2_2** | **B2_3** | **B2_4** | **B2_5** | **B2_6** | **B2_7** | **B2_8** | **B2_9** | **B2_10** |
| Level 0 | .82 | .58 | .33 | .63 | .69 | .71 | .67 | .53 | .50 | .54 |
| Level 1 | .06 | .08 | .11 | .06 | .07 | .03 | .02 | .05 | .07 | .02 |
| Level 2 | .02 | .17 | .07 | .02 | .01 | .06 | .09 | .15 | .07 | .06 |
| Level 3 | .09 | .08 | .15 | .18 | .15 | .18 | .15 | .22 | .10 | .25 |
| Level 4 | .02 | .08 | .33 | .10 | .07 | .03 | .07 | .05 | .26 | .13 |